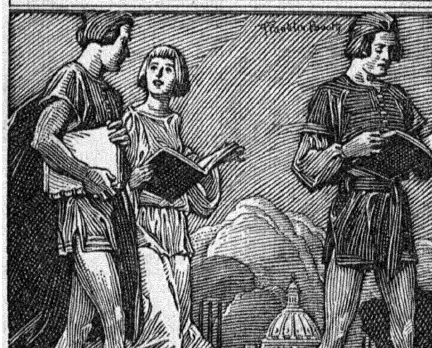
DAMAGE BOOK



PSYCHOLOGY

A Factual Textbook

BY

EDWIN GARRIGUES BORING
Harvard University

HERBERT SIDNEY LANGFELD
Princeton University

HARRY PORTER WELD
Cornell University

AND

COLLABORATORS

NEW YORK
JOHN WILEY & SONS, Inc.
London: CHAPMAN & HALL, Limited
1935

PRINTED IN THE UNITED STATES OF AMERICA

THE HADDON CRAFTSMEN, INC.
CAMDEN, N. J.

CONTRIBUTORS

John Gilbert Beebe-Center
Edwin Garrigues Boring
Charles William Bray
Warner Brown
Leonard Carmichael
Dwight Westley Chapman
Karl M. Dallenbach
Harry Reginald DeSilva
Shammai Feldman
George Humphrey
Daniel Katz
Carney Landis
Herbert Sidney Langfeld
John Alexander McGeoch
Donald McLean Purdy
Miles Albert Tinker
Harry Porter Weld
Ernest Glen Wever
Michael Jacob Zigler

PREFACE

Experimental psychology has reached the stage of maturity. There is a vast amount of well-substantiated fact which forms the foundation of the science, and we have felt that these facts should be presented to the young student of psychology in terms free from the bias of metaphysical presuppositions or of psychological systems. In short, our aim has been to present such a factual text as one should expect from a science. Accordingly, theoretical discussions have been to a great extent omitted and controversial points avoided. Our purpose has been to achieve, not a handbook encumbered with a mass of detailed information, dates, the names of investigators and the titles of monographs, but a generalized statement of fundamental facts in so far as generalization is possible at this time.

Even a cursory examination of the book will reveal an uneven balance among the various chapters. This fault, if it is a fault, is that of psychology and not of the authors. Psychological knowledge is much more extended and precise in some fields than in others. We believe that it is better to present the facts as they exist rather than to distort the picture, by suppression in some chapters and overemphasis in others, in order to achieve proper proportion.

Research has been intense and extensive in psychology during the last few decades, and it was evident to us that it is no longer possible for one man correctly to evaluate and critically to select the results of experimentation over more than a limited range. We have therefore asked experimentalists to prepare chapters in the fields in which they are specialists. Much of the material is printed substantially as it was submitted by these collaborators. We have for the most part merely suggested revision by the authors, or have cut, or have rephrased in order to avoid overlapping and to obtain, so far as possible,

a unity of style and a balanced form. In all cases, however, we must assume final responsibility.

We take this opportunity to express our deep appreciation of the hearty and enthusiastic cooperation of the collaborators who have at all times in the preparation of this book generously acceded to our plans and wishes.

E. G. B.
H. S. L.
H. P. W.

CONTENTS

Chapter I

THE NATURE OF PSYCHOLOGY

A person is a living organism acting in relation to its environment. This environment is the world at large, and it is always changing. Some of these changes affect the organism, and often, when the organism is affected, we find that its behavior is altered. Such a change in behavior may in turn produce change in the environment. So the process keeps up, constituting what is commonly called 'life.'

The organism in its environment. Let us examine such a living person from a scientific point of view. Then we find that we have to do with an extremely complex physical system: an organism and its immediate environment. Events within this system have a definite causal order, thus: a stimulus affects a sense organ, something happens in the nervous system and the organism does something. It is in the understanding of this system that psychology is primarily interested.

1. A *stimulus* is such a change of energy in the environment as affects a sense organ. Light becomes a stimulus when it enters the eye and falls upon a sensitive retina. Sound becomes a stimulus when it gets through to the inner ear where the auditory nerve endings lie. A weight upon the skin, a cough-drop in the mouth, a perfume in the nostrils—these are stimuli provided they actually do stimulate. They are not stimuli if the sense organ is in any way prevented from functioning. A distant object may be called a stimulus if it gives rise to light or sound or any other form of energy that presently stimulates the organism. Most often, when we speak of a stimulus, we are thinking of some object that is under control in an experiment. Thus a tuning-fork is a stimulus to tone because by it we can control the sound that reaches the ear.

Sometimes it is preferable to call this external cause of action a *situation.* A situation is thought of as more complex than a stimulus and also as depending for its effect upon the particular past experience of the person. Red light falling upon the retina is a stimulus, but a traffic officer blowing his whistle and holding up his hand is a situation. The distinction is one of convenience and is not sharp.

2. A stimulus or situation affects an organism through its sense organs. The sense organs respond in their own peculiar ways to the energies that stimulate them. Specific nerve impulses result, and they pass with many modifications into the spinal cord and the brain. There the incoming impulses determine outgoing impulses that arouse muscular movement or the secretion of glands. This second large phase of the total system which we are considering is *neural.*

The outgoing impulses are determined only in part by the impulses that the stimulus or situation has aroused. Action depends more upon the constitution of the nervous system than upon the stimulus. A whistle produces this or that reaction of the organism, according to its meaning, that is to say, according as the nervous system because of past learning directs the effect of the whistle into one channel of action or another.

Not all stimulation of the sense organs gives rise to action. Many incoming impulses seem to get lost. However, we can never know very much about such impulses, because it is only those nervous events that lead to action that we can study easily. When a man is said to *see* that a card is red, we mean, usually, that he *says* that the card looks red to him, or else that he *makes a signal* to show that the card is red. He must *do* something about the stimulus if we are ever to find out that the stimulus was effective, and it is for this reason that we are not far wrong in saying that psychology studies only those nervous events which lead, or can lead, to action.

3. The last phase of the total system is *response* or *behavior.* We use the term *response* when the result is simple and scientifically definite, like the movement of a finger in response

to a signal, the contraction of the pupil of the eye in response to a bright light or the secretion of saliva in response to food in the mouth. Complex systems of responses which depend upon the meanings involved are better called *behavior.* To escape from a dog or to defend a client at court is behavior. *Response* is properly correlated with *stimulus*; *behavior,* with *situation.*

The person or animal who responds to a stimulus or behaves in a situation is called by the psychologist the *subject.* The pages of this book are devoted to the study of this living organic subject.

4. It must be noted that the chief interest in the subject is his *differential response to stimulation.* The only way in which we can tell that a subject sees red as different from green is that he responds differently to it. In the same way we can distinguish a subject's friends from his enemies. In these cases the subject may be a person, a dog or some other animal. Sometimes the situation is greatly involved, as when the response is the mathematician's insight into the solution of a difficult problem. Yet in every case the response is significant because it is differential, because it discriminates the stimulus or situation, past or present, from something else that it is not. Red has particular identity because the response to it can be specific, differentiating it from everything else.

This fact of the differentiating nature of response and its significance for psychology appears with especial clarity in animal psychology. Can a rat perceive spatial form? We know the answer when we know that the rat can respond differently to a triangle and a square. Can a raccoon understand what *three* is? We know the answer if we know that the raccoon can go three times around a block in a maze and then turn away, for it is then that the subject is discriminating the third trip around the block from the others. There is no problem about the existence of conscious experience in the human subject that cannot be paralleled by a problem of discrimination in the animal subject, and some psychologists have even gone so far as to say that consciousness *is* discrimination.

Phenomenal experience. It is the oldest and most deeply rooted tradition in psychology that there is a fundamental difference between mind and matter, and thus between conscious experience and the living body to which the experience belongs. For most of its life scientific psychology has been distinguished from physiology on the ground that psychology deals primarily with the mind and only secondarily with the material body upon which the mind depends. Nowadays this distinction is breaking down because in experimental psychology one observes the mind by observing the discriminatory responses or the verbal behavior of the subject.

As a practical matter it can be made obvious to everyone that personal experience can be observed immediately. If one holds a pencil up before the mirror the immediate experience is that of two pencils. Here we have a datum of mind or consciousness. The method of physics tells a different tale; there is only one pencil, although there is also its reflection. If, after looking at the setting sun, I close my eyes and continue to see a brilliant disk, then that after-effect is also mental. There is no longer a bright light falling on the retina, and, if we could know what the neural processes are like, they would not seem 'bright.' In other words, we can observe experience directly.

The term for this direct observation of experience is *introspection*. The data that introspection shows are called *phenomena*, and the totality of them is *phenomenal experience*. Phenomena are such as memories, thoughts, dreams, perceptions, sensations. Illusions always point to phenomena with especial clarity because in them the lack of correspondence between the phenomenon and its stimulus is notable. The straight stick (stimulus) looks bent (phenomenon) in water. The steady room (stimulus) spins around (phenomenon) for the dizzy man. The common pedestrian (stimulus) looks sinister (phenomenon) to the paranoid patient.

When the direct observation of phenomenal experience is being employed in psychological experimentation, the subject is often called the *observer*, because the subject takes over for the time being the observational responsibility from the ex-

perimenter. The occurrence of the term *observer* in this sense gives notice that the introspective method is in use. An animal is never called an observer, because it cannot assume scientific responsibility for the experiment, even though an animal in making a discrimination between two colors might be said to have 'observed' the difference.

There have been many theories of the relation of phenomena to events in the brain. The theory of *interaction* is that physiological events affect mental events, and conversely. This theory has not been popular because it appears that the physiological events have their own nervous causes and effects and thus form a closed system. This difficulty is avoided, however, in the theory of *psychophysical parallelism*, which supposes that the mental events, the phenomena, merely parallel processes in the brain, which are themselves links in a series of causes and effects that runs continuously from stimulus to response. Others pose a *double-aspect* theory, holding that the events in the mind and the brain are not different entities but different aspects of the same underlying entity. Still others support an *identity* theory on the ground that the underlying entity is not known and that the two aspects are therefore only the same event observed by different methods.

The student can see that there is only a shade of difference between these theories when they are considered in this order. Moreover the difference, being speculative, cannot affect research one way or the other. The first two theories are *dualistic*; the last two are *monistic*. Yet the introspective method is available to either kind of psychologist. The dualist uses it, describing phenomena for their own sake and relating them to neural events as best he may. The monist uses it, regards the results as implying something about the brain, and looks for the same neural relations as best he may. The monist employs the method of sensory discrimination with animals and thinks he has got in his result a measure of neural capacity. The dualist uses the same method and thinks that the same result indicates the differentiation among phenomena in the animal's consciousness. The distinction is not really vital, and for that

reason it has been possible for the present volume to be written by authors who have not committed themselves to a definite theory of the relation of mental phenomena to events in the brain.

Schools of psychology. The emergence and subsequent decline of different schools of thought has been one of the most striking characteristics of the history of modern psychology. The schools, however, while promoting much controversy, have not in general changed the face of experimental psychology, although they have favored one feature for a time at the expense of some other. We could afford to neglect them entirely were it not for the fact that certain historical trends in psychology have definitely affected the experimental work, and that nowadays two words—*behaviorism* and *Gestalt*—find their way into magazines, newspapers and the radio, so that they need explicit definition here.

1. The *classical tradition* in psychology was dualistic, and its particular theory of mind and body was mostly psychophysical parallelism. Wilhelm Wundt (1832-1920), sometimes called the founder of experimental psychology, took that view from philosophy and established it firmly in the new science. From 1860 until well into the present century, psychology considered that its main business was the study of phenomenal experience by the method of direct observation. The physics of the stimulus and the physiology of the nervous system were regarded as subjects of secondary importance, to be studied because of their relation to the phenomena. The age and respectability of this tradition are such that the interest in phenomena still dominates a great part of experimental psychology, and physiological psychology is still thought of by some persons as not quite truly psychology.

2. *Behaviorism* was founded by John B. Watson in America in 1913. It was a protest against the classical tradition. Watson maintained that the method of discrimination yields with animals all the results that the method of introspection gives with persons. He thus formulated the psychology of stimulus-and-

Chapter 2

THE RESPONSE MECHANISM

A knowledge of the structure and function of the living organism is important for a complete understanding of the conditions under which behavior and conscious experience occur. For most of the purposes of scientific psychology, however, what may be termed the *response mechanism* may be studied without specific reference to the other systems of the living body, such, for example, as those concerned in the circulation of the blood and in digestion.

The response mechanism in man and in the higher animals includes: the sense organs or *receptors*, which are organized to receive stimulation and to start processes of excitation in the living individual; the *nervous system*, which is specialized for the propagation of excitation; and the muscles and glands or *effectors*, which are specially developed to make reaction possible. It is largely by means of this highly complex total mechanism that the living organism itself is given a functional and integrated unity, and that in turn the integrated individual responds to those physical energies of the environment which effectively stimulate it.

DEVELOPMENT OF THE RESPONSE MECHANISM

The evolution of the response mechanism in the animal series. This responsive relationship between the adult living human being and the physical energies of his environment may be made more clear by a consideration of the gradual evolution of the response mechanism. Even the simplest organisms respond to environmental energies. A

This chapter was written by Leonard Carmichael of Brown University.

unicellular animal, such as an amoeba, has obviously no specialized cell complexes or organs for the reception of stimuli, the transmission of excitation or the effecting of response. Yet the amoeba may be considered to be a self-contained, living system of energies which may be acted upon and dynamically changed by many of the same physical stimuli which are significant in determining the behavior adjustments and indeed, as will be shown in later chapters, the whole psychological life of man. Radiant energy, vibrations in material media, chemical and other energies act upon the amoeba and initiate processes in its single cell. As a result of the processes so initiated, the total orientation of the cell in relation to its environment may be changed. Thus the amoeba responds to its environment (Fig. 1, I).

In the evolutionary development of the multicellular organisms, the progressive improvement of the response mechanism may be considered as involving a number of clearly marked stages. First came the specialization of effector cells without special receptors or central nervous system. These effectors were primitive muscle cells which respond when directly stimulated. This stage is exemplified in the so-called *independent effectors* in the sponge (Fig. 1, II). Next came the establishment of specialized receptors related to effectors by a diffuse *nerve-net* system, a *receptor-effector* system seen typically in the sea-anemone (Fig. 1, III). Finally, in forms such as the worms, came the development of a true central nervous or *receptor-adjustor-effector system*, as it has been called (Fig. 1, IV and V). The receptor-adjustor-effector mechanism is seen in increasing effectiveness in the series of mammals and in man.

The adjustor mechanism makes possible the different connections between incoming and outgoing impulses. The possibility of this switchboard-like action is in part related to the fact that the continuity of the nervous system, as seen in the old nerve-net stage, has given place in the receptor-adjustor-effector system to relatively independent nerve cells or *neurons*. In understanding the functions of the human nervous

system a clear knowledge of the structure, function, independence and, indeed, the interdependence of neurons is important (see Fig. 2).

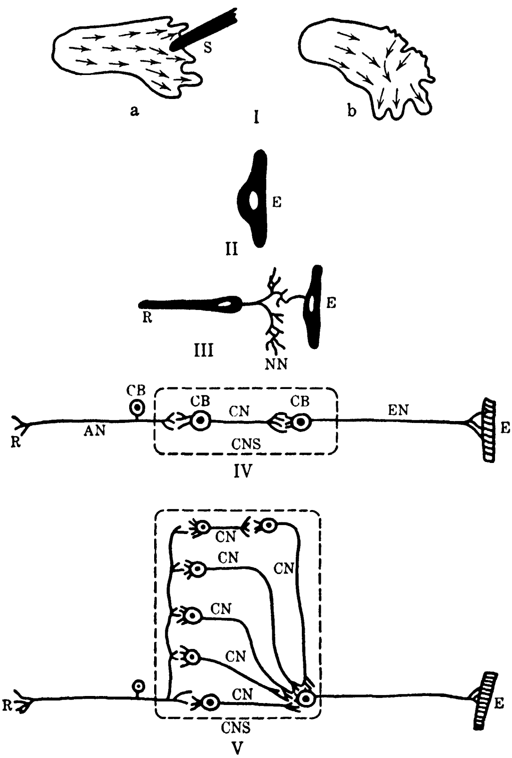


FIG. 1. STAGES IN PHYLOGENETIC DEVELOPMENT OF THE RESPONSE MECHANISM. I. Amoeba. (*a*) Just stimulated by glass rod (*S*); (*b*) change of flow of protoplasm and response of amoeba to such stimulation. II. Independent effector cell. III. Receptor-effector mechanism, with some indication of nerve net (*NN*). IV. Simple receptor-adjustor-effector mechanism or reflex arc. V More complex receptor-adjustor-effector mechanism. Abbreviations. *S* = stimulus, *E* = effector, *R* = receptor, *NN* = nerve net, *AN* = afferent neuron, *CB* = cell body of neuron, *CN* = connector neuron, *CNS* = central nervous system, *EN* = efferent neuron.

Among the many characteristics which distinguish the behavior of an organism possessing a central nervous system from one with a less adequate response mechanism is the ability of the organism to learn. Lower forms have some hint of this

ability, but in its developed form learning is a function of the adjustor mechanism. There could be nothing like adult human behavior or adult human consciousness, did the human being not possess this capacity. The development of this characteristic of relatively permanent modifiability of the *central ganglia* or central nervous system, that is to say, the capacity to learn, has made possible the human being's unique independence of his immediate environment.

The evolution of the response mechanism in each individual. Every adult living individual begins life as a single fertilized germ cell. During the early days of prenatal life other cells develop from this single cell. These cells come eventually to form tissues and organs, which pass through an elaborate series of changes and eventually constitute the recognizable human body.

During the very early part of human prenatal development structural changes begin which are fundamental to the formation of the central nervous system, such great receptor organs as the eye and ear and the effector apparatus. At length, after a most remarkable series of alterations, the sense organs, nervous system and muscles of the yet unborn child begin to work so that the stimulation of receptors leads to effector response. Typically at least five months before normal birth such true responses begin. From this time on until birth, patterns of response, as released by internal and such external stimuli as are present, develop. The outcome of this course of change in prenatal behavior is that the young organism is increasingly able to respond adaptively to the stimuli of its environment. That is to say, as the normal development of the response mechanism, and especially the adjustor mechanism, goes on in the human infant, more and more precise and discriminative responses to the physical energies of the environment become possible. What is usually called *growth* and what is usually called *learning* cooperate in bringing about those changes in the response mechanism which underlie this development.

Thus, both in the animal series and in the individual, the story of the development of the response mechanism is the story of the evolution of a greater and greater capacity on the part of the organism to respond *differentially* and, from the standpoint of the welfare of the organism, *adaptively* to the physical energies of the external world.

The role of the receptors in the response mechanism. At all stages in the development of the organism effective stimuli may be described as physical energies. In the higher organic forms such energies usually act on specialized receptors. Some of these receptors are at the surface of the body and are located so that they may be easily affected by external environmental forces. These are called *exteroceptors* and are exemplified by the receptor cells of the eye. Some receptors are imbedded in the bodily substance itself. Typical of such receptor cells are the sensory cells of the muscles, which are stimulated by the movement of the muscle substance. Such receptors are called *proprioceptors.* There are also receptors associated with the lining of the digestive tract, and these are sometimes characterized as *interoceptors* (see Fig. 8). The exteroceptors, proprioceptors and interoceptors alike are associated with the peripheral endings of *afferent* neuron fibers. These fibers pass from the periphery to the central nervous system. From the central nervous system *efferent* neurons in turn pass out to effectors. This total path from receptor to effector, involving all of the great divisions of the response mechanism, is called a *reflex* or *response arc.* (Two diagrammatic forms of such arcs are shown in Fig. 1, IV and V.)

The total living organism may be conceived of not only as an anatomical entity but also as a complex system of energies. The essential process of receptor action may be described in terms of the manner in which an external physical energy is able to bring about a change in specific subsystems of energy in the organism. Stimulation thus initiates processes in the animal which typically lead, at least eventually, to effector response and a change of orientation of the individual in rela-

tion to its environment. Stimulation is in some respects analogous to the finger pressure on the trigger which initiates the release of energy in the gunpowder in a cartridge, and thus leads to the expulsion of a bullet from a gun. Obviously in the

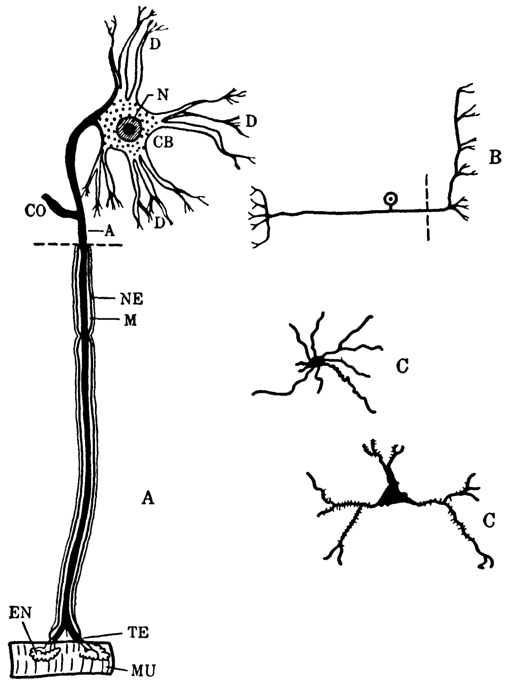


FIG. 2. STRUCTURE OF SOME TYPICAL NEURONS.

A. A typical efferent (motor) neuron. *B*. A typical afferent (sensory) neuron (in less detail than *A*) *C*. Typical central (connector) neurons (in less detail than *A*). Abbreviations *D* = dendrites, *N* = nucleus, *CB* = cell body, *CO* = collateral, *A* = axon, *NE* = neurilemma sheath, *M* = myelin (medullary sheath), *MU* = muscle, *EN* = motor end plate, *TE* = terminal arborization or end brush.

cartridge the explosion of the stored energy, and not the movement of the finger, drives the bullet. The release (by stimulation) of energy stored in the receptor does not propel matter through space, but it does truly initiate *progressive release of energy*. That is to say, the physical energy of the stimulus does not itself enter the receptor; it merely sets off cer-

tain energies of the organism located in the receptor, and then other progressive releases of energy follow.

In the higher animals not all receptors are alike in the work they do. The main function of a specialized receptor is to be easily 'set off' by one sort of physical energy and to be with difficulty or not at all stimulated by any other form of energy. Technically the energy for which a receptor is specialized is called, in relation to that receptor, *adequate.* All other energies, even though they may set the receptor in action, are called *inadequate.* There is, however, reason to suppose that the end result of receptor stimulation, *i.e.,* the initiation of a *nerve impulse,* is in each case similar in certain respects. For example, if radiant energy from the so-called visible spectrum stimulates the receptors of the retina, we have adequate stimulation of the eye, and processes are started which are correlated with the experience of light. But pressure on the eye may inadequately stimulate the receptors of the retina. Once the retina has been so stimulated, however, the processes which result lead to activities which are correlated with the experience of light and not of pressure.

GENERAL STRUCTURE AND FUNCTION OF THE NERVOUS SYSTEM

We have just seen that, as a result of the stimulation of a receptor by external physical energies, certain energy changes are initiated in the receptor. This energy change involves a spread of a chemical and electrical disturbance in the living cell which is activated. In the simplest cases in the human body the receptor is itself merely the free ending of an *afferent neuron of the peripheral nervous system.* Usually the receptor is a specialized cell associated with such a neuron. The afferent peripheral neuron itself is typically a continuous protoplasmic thread connecting a receptor with the neurons of the central nervous system. The peripheral fiber of a single neuron may thus be several feet long, for it is unbroken from receptor to central nervous system, but it is microscopic in diameter. In most cases each such neuron fiber is insulated by

special sheaths. A great many such insulated fibers are ordinarily held together by non-neural tissue to form a *peripheral nerve*. Such nerves usually contain, at least for certain distances, many independent fibers, of which some are afferent and others efferent.

The chemical and electrical disturbances which are set up in the receptors as a result of stimulation in turn initiate similar processes in the protoplasmic threads of the afferent neurons. These 'disturbances' spread along the fibers. Such *propagated regions of increased activity* are called *nervous impulses*. From the standpoint of a knowledge of the response mechanism as a whole, an understanding of the nature of these impulses is most important.

The nature of the neural excitation. When the stimulated receptor starts a process of activity in the afferent peripheral neuron, the receptor does not pour something into the neuron. Instead, by a trigger mechanism, the energy of the active receptor upsets the energy balance of the neuron and thus in turn initiates a further disturbance in the afferent fiber. This disturbance is then propagated along the axon. Experiment shows that, if a single neuron is excited in this way at all, it is excited to its maximum extent. This generalization is known as *the all-or-none law*. This law may be stated more formally as follows: The magnitude of the activity in any single neural functional unit is as great as it can be in that unit at that time and is independent of the magnitude of the energy exciting it, provided only that the stimulating energy is sufficiently strong to excite the neuron at all. A crude analogue of this characteristic of nerve activity may be found in a burning trail of gunpowder, or indeed a simple burning string. The energy which is released and the characteristics of the moving flame are not dependent on the quantity of energy of the match that ignites the train, but rather the combustion is locally determined as each point in the train is successively burned. Careful experimental work has shown that the characteristics of the nervous impulse at any point on a nerve

fiber are determined by that part of the fiber, and not by the energy which in the first place started the impulse.

One fundamental change which may be studied in determining the nature of the propagated disturbance in the neuron is an electrical phenomenon, the *action potential* or *action current.* The active region of the neuron fiber is electrically negative in relation to the unexcited portion of the same fiber. Fig. 3 shows how such a propagated region of negativity may be recorded on a galvanometer. This impulse, or region of excitation, travels in mammalian nerve at a speed of approximately 100 meters a second. Such a speed, though relatively fast, is, of course, in no way comparable to the speed of light or to the speed of an electric impulse in a wire.

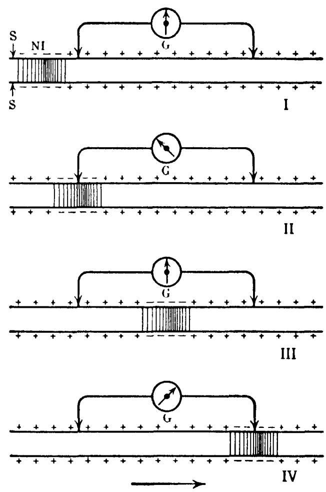


FIG. 3 PROPAGATION OF AN ELECTRICAL DISTURBANCE ALONG A NEURON FIBER.

I, II, III, IV show successive time intervals as the impulse passes from left to right The galvanometer deflection is indicated in each case It will be noticed that the impulse is marked by a negative deflection Abbreviations: S = stimulus, NI = nerve impulse, G = galvanometer.

If we return to the analogue of the burning trail of gunpowder, we may say that, once the gunpowder trail has been burned, it cannot be ignited again until new energy in the form of a new trail of gunpowder has once more been laid. In the nerve, careful study of time relations in stimulation has shown a process which is roughly comparable to this need for replenishing combustible material before re-ignition can occur. It has been shown that there is a period immediately following the peak of activity in a neuron during which it cannot be activated again, no matter how strong the stimulus may be. This time interval is technically known as the *absolute refractory period.* Following this period there is an inter-

val during which the neuron may be stimulated, provided the stimulus is stronger than the minimal stimulus which would have been effective in starting an impulse in that neuron when it was normal or resting. This interval is called the *relative refractory period.* Immediately after the relative refractory period the neuron may be thought of as having recovered, and thus as being ready again for activation by a stimulus of normal magnitude. It has indeed been demonstrated that, under certain circumstances, immediately following the

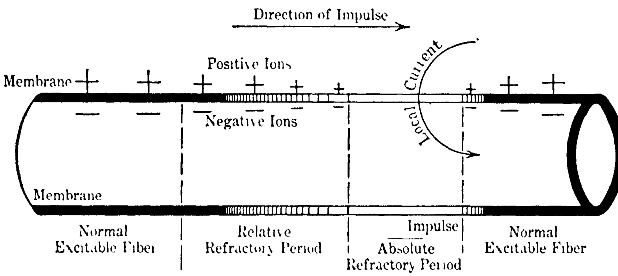


FIG. 4. MEMBRANE THEORY OF NERVE CONDUCTION.

Hypothetical and schematic The semipermeable membrane is shown in black with the positive ions on the outside and the negative ions on the inside. The impulse is traveling from left to right It consists of a local current as the positive and negative ions unite when the membrane becomes permeable, and this current renders permeable the membrane ahead of it so that the impulse continues on. The permeable membrane is shown in white, and the membrane, which is being restored after the passage of the impulse, is shaded. Thus the diagram also shows the refractory periods. Adapted from Boring.

relative refractory period, a stimulus of an intensity lower than that required to excite the resting nerve may be effective. This supernormal period, however, is possibly an artifact of the laboratory. Characteristic time relations of these processes are shown in Fig. 4. Careful experimental work has also demonstrated that the events involved in the passage of a nerve impulse produce a measurable amount of heat. It has further been shown that oxygen is used up and carbon dioxide given off in this activity.

A short intense stimulus has been found to set up processes in the receptor which outlast the short duration of the stimulus. The time of the cessation of this effect has been found

to be related to the intensity of the exciting stimulus. Thus, nerve impulses which are *periodic* in character, because of the refractory periods of the tissues involved, may follow for a short time from an excited receptor after the receptor itself is no longer being activated by an outside energy.

Experimental study of the relative refractory periods of the receptor and the individual fiber has also shown that it is possible for an intense stimulus to excite such a mechanism into a relatively rapid series of successive impulses, while a weak stimulus acting upon the same mechanism may produce a less rapid series. The succession of discharges in a peripheral fiber becomes greater the more intense the physical energy of the stimulus applied to its receptor. The total limits of this frequency are, it can be seen, always determined by the time limits of the relative and absolute refractory periods of the activated cells in question. This increase in frequency of discharge in relation to the increased intensity of stimulation has been shown to occur experimentally where but a single receptor and a single neuron are involved.

In any consideration of the correlation between the stimulus and the nervous impulse, it must never be forgotten that in many sense fields the receptors present a surface of 'receptive points.' In excitation by a physical stimulus having area, or in certain cases, it may be, by the spread of excitation from a strong single point of stimulation, a number of fibers may be activated. In some receptor fields, such as the retina of the eye, the relationship between receptors and associated neurons is most complicated. The complications in the retina, which is really embryologically part of the central nervous system, cannot be considered in detail here. It is altogether probable, however, that the various sensory cells in any typical receptor field may *differ* with respect to the *lower thresholds of excitation.* That is to say, a stimulus of a particular physical intensity may call into action some, but not all, such receptors. In a condition of this sort, it might well be that the boundary and, indeed, the surface area of the non-punctate stimulus would be approximately represented by the excited receptors, without

necessarily involving the activation of all the receptors within the area covered by the stimulus. The number of receptors involved in any given stimulated area at a specific degree of stimulus intensity has been spoken of as the *density* of stimulation.

It may also be noted that not all peripheral fibers are alike in the characteristics which determine the varying lengths of absolute and relative refractory periods. Thus, from fiber to fiber there is a possibility of different frequencies of discharge. It has been demonstrated, for example, that, in general, the larger the diameter of the neuron fiber the higher the possible rate of frequency of discharge. Marked variations in the diameter of fibers have been shown by microscopic study. Anatomical differences between neurons may also be correlated with differences in the *shape* of the rise and fall of the region of activity in a propagated disturbance, as well as with the speed of conduction of this disturbance.

None of these facts related to the multiple-receptor and multiple-fiber mechanism must be forgotten in working out the relationship between the measured characteristics of the physical stimulus and the physiological events of the activated nervous system. For example, from what has just been said it will be obvious that with stimuli which affect an areal receptor surface it may be possible that, as the intensity of the stimulus increases, an increasing number of individual receptors will be stimulated, each fiber reacting in an all-or-none manner. Intense stimulation may also lead to more frequent impulses because it can reactivate the neuron before the recovery process of the relative refractory phase is complete. It thus seems that an increase in the intensity in stimuli may be correlated, in the peripheral nervous system, with an increase in *number of units* affected and also with an increase in the *number of impulses* per second in each fiber involved. The part played by these two factors in determining the neural correlate of stimulus intensity is considered later (pp. 193-195).

It has become clear that the characteristics of the propagated excitation which occurs in the fibers composing a peripheral

nerve may differ in a number of ways. Of these ways the following seem to be most important: (1) The frequency of impulse may vary in any given fiber. (2) The frequency of impulse may vary from fiber to fiber. (3) The total number of fibers excited may vary as the total area of excitation in the receptor field is changed. (4) In two equal areas the total number of fibers excited may vary in relation to the strength of stimulation. (5) The duration of activity may vary. (6) The form and speed of what may be called the energy gradient of the active portion of the neuron, the 'impulse,' may vary from neuron system to neuron system.

The most important theory that has been developed to account for the chemical and physical processes underlying neuron excitation is the *membrane hypothesis of nerve conduction.* According to this view each threadlike fiber of a neuron is physiologically as well as anatomically complex. In its resting state such a fiber may be thought of as a cylinder, the walls of which consist of a semipermeable membrane. The wall is called semipermeable because it forms a differential barrier to the passage of chemical or electrical elements. In the living neuron this membrane is electrically polarized: positive ions are on its outside and negative ions are on its inside surface. According to this theory the action current is a local change in this polarization. The action current is therefore related to the establishment, by a process of depolarization, of a new electrical potential difference on the outer surface of the membrane between an inactive and positive region and the now active and negative region. This electric change causes a local reorganization of ions on the surface, involving also to some extent the membrane ahead of it. In this way the part just beyond the excited region in turn becomes permeable and thus a spread of depolarization is brought about. This spread may occur in either direction in the peripheral fiber, but once established it can move in but one way. This directionality is due to the fact that the disturbance of the ions accompanying activity is such as to require time, as explained above, for an

elaborate process of restoration of resting polarity in the recently active region (see Fig. 4).

We have now considered something of the nature of the activities in the organism which result from external stimulation as they appear in the receptor and in the peripheral nerve. Next we must follow the progressive release of energy into the central nervous system and consider the special phenomena which are related to the spread of such excitation in the brain and spinal cord.

The structure of the central nervous system. The central nervous system in man is made up of the spinal cord and the brain (Fig. 5). The spinal cord is the part of the nervous system that is enclosed in the jointed bony case of the vertebral column. It is connected with receptors and effectors by more than thirty pairs of spinal peripheral nerves. The spinal cord is primarily to be thought of as a cable, by means of which impulses initiated at the receptors of the body may be transmitted to and from the higher centers of the brain. The cord is also in its own right a center for the correlation and adjustment of simple reflex arcs.

Continuous with the spinal cord and protected by the bony case of the skull is a very complex system of nerve centers and communicating tracts known as the brain. Immediately above the cord and in continuity with it is the *medulla oblongata.* Like the cord the medulla is an important adjustment center in its own right, but it is primarily, also like the cord, to be thought of as a great cable of fibers connecting the spinal system below with the higher brain centers above. Situated dorsally to the medulla and, as it were, off the main track of the central nervous system are the two hemispheres of the *cerebellum,* which functions in the coordination of bodily movements. Below the cerebellum is a large structure, the *pons,* which is made up of fiber tracts and specialized adjustment centers. Above the cerebellum and pons is an elaborate series of special connecting centers, all of which play an important part in the adjustment of impulses and in the adaptation of the

organism to its environment. Much is known concerning these centers, and much is still to be discovered. It is impossible, how-

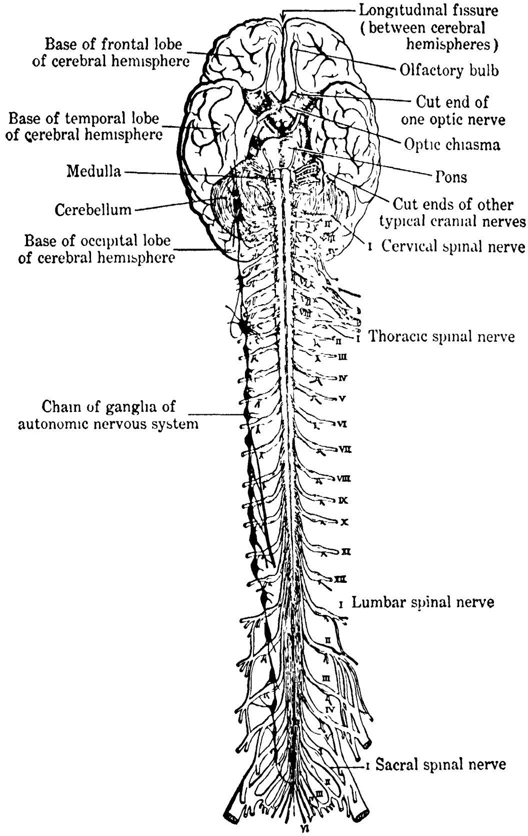


FIG. 5. BASE OF BRAIN AND SPINAL CORD.
Heavy black structure at left of cord indicates part of autonomic nervous system. Adapted from Herrick (5).

ever, to review the anatomical relationships of these centers here. Nevertheless, it is important to note that before we reach the level of the cerebral cortex we pass through a complex group of connecting centers which may be roughly char-

acterized as the *thalamus* or the thalamic region. In a later chapter of this book the significance of this region in emotion (see pp. 394f.) will be treated. The cranial peripheral nerves all enter and leave the brain below the level of the cerebral hemispheres.

In man by far the largest part of the brain is the great *cerebrum*, which is divided into two *cerebral hemispheres.* These structures are large, closely organized masses of neurons which almost fill the skull. They are constructed to receive impulses from lower levels of the central system and to send impulses back to such centers. They form an adjustment center superimposed, as it were, upon the lower, more immediate, connecting centers of the central nervous system.

There has been much speculation about the relationship between sheer brain weight and psychological ability. When a formula is constructed which makes possible the comparison between the ratios of brain and body weight, it is found that there is approximate agreement between the relative increase in brain weight and adaptive ability, so far as the various species in the phylogenetic scale are concerned. That this correlation also holds true statistically in comparing human beings of different intellectual ability has not, however, been demonstrated.

The general functions of the central nervous system. In any consideration of the central nervous system one may well remember that, no matter what the complications of this system may be, it is possible to look at it as basically organized to make connections between incoming and outgoing nerve impulses possible. The mechanism of these connections is, of course, more than a simple transmission. Sherrington, one of the most thorough students of this field, has said of the central nervous system that it is "an organ of reflex reinforcements and interferences, and of refractory phases, and shifts of connective pattern; that it is, in short, an *organ of coordination* in which from a concourse of multitudinous excitations there result orderly acts, reactions adapted to the needs of the organism, and that these reactions occur in arrange-

ments (*patterns*) marked by absence of confusion, and proceed in *sequences* likewise free from confusion" (27, 313).[1]

It has been calculated that there are probably at least twelve thousand million nerve cells in the central nervous system. This inconceivable complexity at first almost seems to balk any hope of understanding central-nervous-system activity. The conception of the response arc, however, provides a key to at

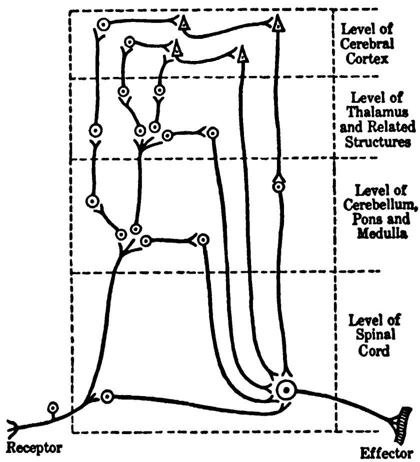


FIG. 6. SCHEMATIC DIAGRAM TO SHOW ALTERNATE LOOPS AT VARIOUS LEVELS OF THE CENTRAL NERVOUS SYSTEM.
Adapted from Bayliss (25).

least some basic central-nervous-system processes. In the central portion of each such arc lie the spinal cord and brain. In regard to that part of the arc that is in the central nervous system, it will be seen from Fig. 6 that some of the paths from a receptor to an effector are relatively short, simple and direct, whereas others are longer and more intricate (see also Fig. 1, V).

Some of the fundamental properties of the central nervous system may be attributed to the *synapses* or points of physio-

[1] The parenthesis is a citation of p. 313 of the reference numbered "27" at the end of this chapter. This system of citation is used throughout this book.

logical junction between neurons in the system of neural arcs. Anatomically it is probable that there is protoplasmic discontinuity at such junctures. This judgment is based upon direct microscopic study and upon the fact that when one neuron dies, the degeneration does not ordinarily pass the point of juncture. On physiological grounds many other characteristics are attributed to the synaptic regions. Such inferences are indirect. Nevertheless, there is evidence that the following special properties of the central nervous system, among others, are in part at least referable to synaptic regions: (1) *Speed.* The rate of transmission of a nervous impulse is slower in the central nervous system than in the peripheral fiber. This delay is attributed to the synapse. (2) *Irreversibility.* Nerve impulses travel in only one direction in the total neural arc. This one-way conduction is attributed to the valve-like action of the synapses. (3) *Susceptibility to drugs.* The central nervous system with its many synapses is much more susceptible to drugs than is the peripheral nervous system. This is true both of the abolishing of functional activity by anesthetics and the heightening of such activity by strychnine. (4) *Variability of physiological action.* This variability, as seen in *reinforcement* (the augmenting of one process by another), in *inhibition* (the partial or complete extinguishing of one functional activity by another) and in *summation* (the cumulative effect of repeated action), may be thought of as brought about in the central nervous system at its synaptic junctures. (5) *Fatigue.* The fact that the central nervous system is more subject to fatigue on continued activity than is the peripheral system may be attributed to changes that take place in the synapses.

There are certain general problems of the central nervous system in which the student of mental phenomena must always be most interested. The first of these is the question of the alleged localization of various psychological functions in the cerebral hemispheres, and the second is the problem of the anatomical and physiological basis of learning. In order to deal with these problems we must consider the methods by

which the detailed function of the brain may be studied. Among those which have been employed to work out a correlation between structure and function in the brain may be mentioned the following:

(1) The method of experimental destruction of tissue in the central nervous system, that is to say, of extirpation, or ablation, in relation to behavior and experience. (2) The method of studying accidental destruction of tissue in the central nervous system in relation to behavior and experience. (3) The method of studying pathological changes in the central nervous system in relation to behavior and experience. (4) The method of direct electrical stimulation of the brain in relation to behavior and experience. (5) The method of local application of drugs on the cortex in relation to behavior. (6) The methods of the histology and embryology of the brain in relation to known facts of behavior and experience. (7) The general physiological methods of brain study in relation to behavior and experience. Changes of brain volume may be studied in relation to various behavior characteristics of the individual. Action currents may be led off from the intact brain or from various areas of the brain and from subcortical centers and an effort made to correlate the characteristics of these currents with other specific activities. Heat production, supposedly concomitant with specific cerebral activity, may possibly also be measured.

As a result of experimental studies of the type outlined above, an increasingly precise knowledge in regard to the correlation between the structures and the functions of various parts of the central nervous system has been secured. In many instances, however, the enthusiasm of investigators, the uncontrollability of clinical observations and the tendency of textbook writers to avoid qualifying phrases in summarizing experiments have led to dogmatic assertions in this field which have subsequently been shown to be unfounded.

The history of the changes in opinion in regard to the supposed functions of the cerebral hemispheres and especially the outer layer of these hemispheres, the *cerebral cortex*, is

long and involved. A competent opinion concerning the function of the cortex may possibly best be given in a quotation from Herrick:

> There is unquestionably mosaic localization of certain physiological functions in the human cerebral cortex. The projection centers are definite areas within which specific systems of projection fibers make their cortical connections in switchboard fashion. But neither these centers nor any sector of the intervening associational tissue can be thought of as performing any one of the distinctively higher cortical functions in isolation. . . . The clinicians' maps of the aphasias, etc., represent (in some of the cases) true pictures of vulnerable points of certain associational complexes. [See Fig. 7.] But most of the charts of functional localization of psychological or other complex functions are misleading fictions (6, 249).[2]

The central nervous system and learning. The work of Lashley may be taken as typical of controlled scientific work in this field. This physiological psychologist has trained animals in a variety of different situations to perform definite and quantifiable tasks. He has then removed parts of the brain of the experimently trained animals and noted the changes in behavior which result. He has also studied the process of re-education in animals that have had their previously learned habits interfered with by brain operations. Among the significant conclusions reached by this investigator and his collaborators is the fact that in the rat, at any rate, the ability to learn is, under certain conditions, reduced in a quantitative degree that is roughly proportional to the quantitative amount of cortical tissue removed.

On the basis of this work and much else that cannot be reviewed here, Lashley concludes that the mechanism of integration and learning, though to some extent localized, is not basically dependent upon connections between specific neurons in the cerebral cortex, but rather that integration and

---

[2] Reprinted by permission of the University of Chicago Press.

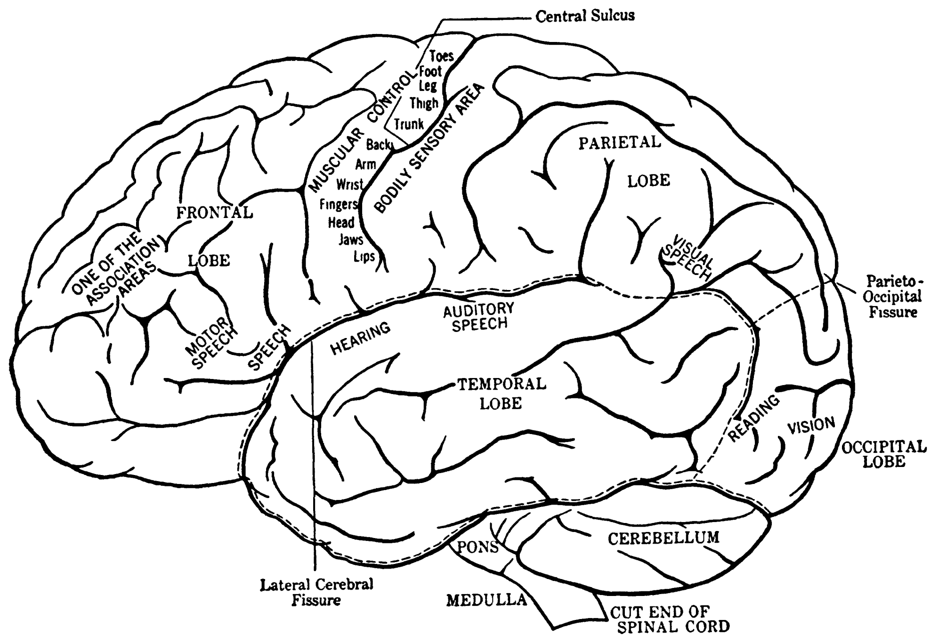


FIG. 7. SCHEMATIC DIAGRAM OF THE LEFT ASPECT OF THE CEREBRAL CORTEX.
Some of the so-called motor and sensory areas shown by labeling. Great caution must be used in interpreting such maps. Adapted from various sources.

learning are a result of dynamic interrelationships between large organized patterns of activity in the neurons of the brain.

Certain neurologists and psychologists, however, have given a slightly different interpretation to some of Lashley's own experimental results and to the findings of other experimental neurologists. It is suggested that subcortical centers, such as those in the thalamus, may be involved in both learning and relearning in animals such as the rat. This hypothesis is more in harmony with the basic conception of the neural arc than is a "mass-action theory" of the brain, such as that suggested by Lashley.

The efferent peripheral nervous system and the effector organs. We have seen that in the central nervous system the results of this sensory peripheral activity may spread over many alternate paths and possibly involve specialized functions characteristic of the total masses of cells in certain parts of the brain. Eventually, as a result of such activity, *efferent* or *motor* neurons must be activated. This central end of the motor neural mechanism may be called the beginning of the *final common path.* This final common path is the sole avenue which all impulses, no matter whence they come, must travel if they are to act on muscle fibers or glands ana bring about response. Thus, activities in various parts of the brain and spinal cord which have resulted, it may be, from exteroceptive stimulation can be brought into relation with impulses from other parts of the central nervous system which have themselves originated, for example, in the receptors of the non-auditory labyrinth and in the proprioceptors of certain muscles. Some of these impulses may mutually strengthen or *facilitate* one another; some may act in such a way as to lead to mutual extinction or *inhibition.* In the normal individual, however, the outcome of such complex activity of adjustment is the finely graded and precisely timed effector response. Thus, activities occur which make up adaptive, intelligent behavior. When psychologists, therefore, use the phrase "stimulus control of behavior" this whole complex series of central-nervous-

system processes must be envisaged, even though the results be presented, as they may often legitimately be, as a quantitative relationship between measured characteristics of the stimulus and measured characteristics of the response.

In a complete consideration of the motor aspects of the response mechanism it would be necessary to deal with the facts of the so-called autonomic nervous system, a motor nervous system which enjoys a measure of independence from the great peripheral and central systems already considered. In this book, however, facts in regard to the autonomic nervous system are presented in the chapter on emotion (see pp. 411ff.).

By means of peripheral efferent fibers and special fibers of the autonomic nervous system, impulses reach effector mechanisms. All muscle, be it *smooth* (that typical of the walls of the intestines), *skeletal* (that typical of the arm and leg muscles) or, like heart muscle, midway between these two types, is specialized for one function. That function is *contraction.* Contraction alone makes the response possible by means of which the stimulated organism *behaves,* that is, readjusts itself in its environment. As we have already seen (pp. 9-11), this property of contraction is the basic function around which all specialization in the neuro-muscular mechanism has been built. (For a diagrammatic schema of the relationship between the various subsystems of the response mechanism see Fig. 8.)

The contraction of typical muscle cells is accompanied by an electrical disturbance which may be recorded and studied by means not unlike those employed in the study of the electrical phenomena of nerve. The electrical changes of the nerve initiate the essential physical and chemical events which lead to the release of stored energy in the muscle and thus to the *work* and *heat* characteristics of activated living muscle.

The secretion of the glands is typically a simple and direct response. Stimulation of a nerve passing to a gland shows that there is a positive relationship between the strength of the stimulus and the amount of secretion of the gland. The complex mechanism by means of which a typical salivary gland may secrete its own weight in saliva in five minutes involves

marked changes in the blood supply to the gland and in the gland itself. These changes are accompanied by electrical phenomena which are probably associated in their inception with the electrical phenomena of the nerve which initiates the activity. Glands may be variously classified. The two most

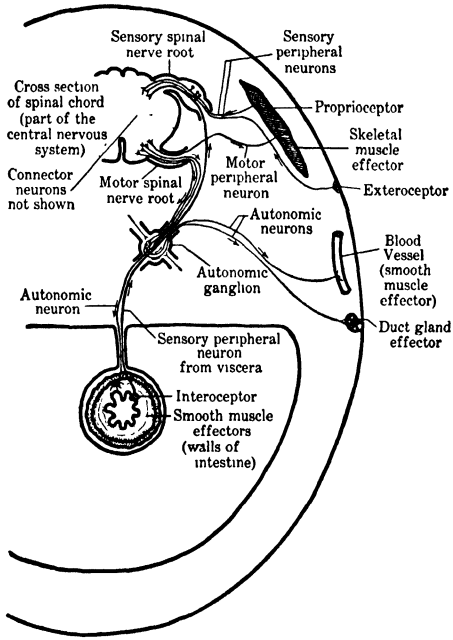


FIG. 8. RELATIONSHIP BETWEEN THE VARIOUS CLASSES OF RECEPTORS, THE NERVOUS SYSTEM AND THE EFFECTORS.

A diagram to show the relationship of exteroceptors, proprioceptors and interoceptors to the peripheral, central and autonomic nervous systems, and to the muscular and glandular effectors of the body. Adapted from various sources.

significant groups are ordinarily thought of as those which pour the product of secretion through a tube into a cavity of the body or out upon the body surface. Such glands are characterized as *duct* or *exocrine* glands. There are other glands in the body, of which the thyroid and the adrenal medulla are typical, which have no duct, but which pour their secretions directly into the blood stream, and are called *endo-*

crine glands. In a complete understanding of the response mechanism the physiological effects of endocrine substances which are carried in the blood stream must be taken into consideration (see pp. 413-416).

Thus by means of effector response the behavior of the living organism is made possible. This behavior is appropriate to the stimulating energies of the environment to the extent that the receptors, effectors and especially the adjustor mechanisms of the organism displaying it are adequate to the task of differential response. In the animal series this capacity becomes more and more effective from amoeba to man. In the life history of the individual human being this capacity becomes more adequate in the long series of changes which occur from the first squirm of the fetus to the adaptive discriminative choices of the adult intelligent man.

REFERENCES

This chapter is elementary. Those persons who wish further reading on the various aspects of the chapter will find the following suggestions of value.

I. On the history of the development of the relationship between psychology and neurology:

1. Boring, E. G. *A history of experimental psychology.* New York: Century Co., 1929.
2. Fulton, J. F. *Selected readings in the history of physiology.* Springfield, Ill.: C. C. Thomas, 1930. Especially pp. 181-295.

II. On the general significance of the response mechanism:

3. Child, C. M. *Physiological foundations of behavior.* New York: Henry Holt and Co., 1924.
4. Crozier, W. J., and Hoagland, H. The study of living organisms. In *Handbook of general experimental psychology.* Worcester, Mass.: Clark University Press, 1934. Pp. 3-23.
5. Herrick, C. J. *An introduction to neurology.* Fifth edition. Philadelphia: W. B. Saunders Co., 1931.

III. On the evolution of the response mechanism in the animal series:

6. Herrick, C. J. *Brains of rats and men.* Chicago: University of Chicago Press, 1926.

7. Jennings, H. S. *Behavior of the lower organisms.* New York: Columbia University Press, 1923.

8. Parker, G. H. *The elementary nervous system.* Philadelphia: J. B. Lippincott Co., 1919.

9. Washburn, M. F. *The animal mind.* Third edition. New York: Macmillan Co., 1926. Especially Chapter 3.

IV. On the evolution of the response mechanism in each individual:

10. Carmichael, L. Origin and prenatal growth of behavior. In *Handbook of child psychology.* Second edition. Worcester, Mass.: Clark University Press, 1933. Pp. 31-159.

11. Child, C. M. *The origin and development of the nervous system from a physiological viewpoint.* Chicago: University of Chicago Press, 1921.

12. Coghill, G. E. *Anatomy and the problem of behaviour.* Cambridge, England: University Press, 1929.

V. On the role of the receptors in the response mechanism:

13. Adrian, E. D. *The basis of sensation.* New York: W. W. Norton and Co., 1928.

14. Hecht, S. Vision. II. The nature of the photoreceptor process. In *Handbook of general experimental psychology.* Worcester, Mass.: Clark University Press, 1934. Pp. 704-712.

15. Troland, L. T. *The principles of psychophysiology.* Vol. II. New York: D. Van Nostrand Co., 1929.

VI. On the general structure and function of the nervous system:

16. Barker, L. F. *The nervous system and its constituent neurones.* New York: D. Appleton and Co., 1899.

17. Forbes, A. The mechanism of reaction. In *Handbook of general experimental psychology.* Worcester, Mass.: Clark University Press, 1934. Pp. 155-203.

18. Lickley, J. D. *The nervous system.* New York: Longmans, Green & Co., 1920.

VII. On the nature of the neural excitation:

19. Adrian, E. D. *The mechanism of nervous action: electrical studies of the neuron.* Philadelphia: University of Pennsylvania Press, 1932.

20. Davis, H. The conduction of the nerve impulse. *Physiol. Rev.,* 1926, 6, 547-595.

21. Hill, A. V., Fenn, W. O., Gerard, R. W., Gasser, H. S. and Parker, G. H. *Physical and chemical changes in nerve during activity.*

Occasional publications of Amer. Assoc. Adv. Sci., 1934. No. 2, pp. 1-29.

22. LILLIE, R. S. *Protoplasmic action and nervous action.* Second edition. Chicago: University of Chicago Press, 1932.

VIII. On the structure of the central nervous system:

23. LAPICQUE, L. Le poids du cerveau et l'intelligence. In *Nouveau traité de psychologie.* Paris: Librairie Félix Alcan, 1930.

24. PAPEZ, J. W. *Comparative neurology.* New York: Thomas Y. Crowell Co., 1929.

See also references 16, 17 and 18 under VI above.

IX. On the general functions of the central nervous system:

25. BAYLISS, W. M. *Principles of general physiology.* Fourth edition. New York: Longmans, Green & Co., 1927. Especially pp. 464-487.

26. CREED, R. S., DENNY-BROWN, D., ECCLES, J. C., LIDDELL, E. G. T., and SHERRINGTON, C. S. *Reflex activity of the spinal cord.* Oxford: Clarendon Press, 1932.

27. SHERRINGTON, C. S. *The integrative action of the nervous system.* New Haven: Yale University Press, 1926.

X. On the central nervous system and learning:

28. LASHLEY, K. S. *Brain mechanisms and intelligence.* Behavior Research Fund Monograph. Chicago: University of Chicago Press, 1929.

See also Herrick (6), especially pp. 178-267.

XI. On the efferent peripheral nervous system and the effector organs:

29. FULTON, J. F. *Muscular contraction and the reflex control of movement.* Baltimore: Williams and Wilkins, 1926.

30. HILL, A. V. *Muscular activity.* Baltimore: Williams and Wilkins, 1926.

On the autonomic nervous system, see:

31. MULLER, L. R. *Lebensnerven und Lebenstriebe.* Berlin: Julius Springer, 1931.

Chapter 3

PSYCHOLOGICAL MEASUREMENT

Every discipline which is in any strict sense scientific finds it indispensable to be able to measure the material which it treats; and psychology is no exception to this generalization.

The usefulnes of measurement in science arises from the fact that methods of measurement are fundamental for the discovery of quantitative relationships; and quantitative, rather than qualitative, relationships are the richest material for the construction of a science.

In order to measure, a science must develop and agree upon certain metrical units and procedures. These units and methods vary in many respects from one science to another, so that we must, in the present instance, devote some attention to those metrical devices which are peculiar to psychology.

Most of the laws of psychology occur in the form of statements of relations holding between (1) some psychological variable of an organism (such as its experience or its behavior), and (2) either the organism physiologically considered or its environment physically considered. Consequently, to put these laws into quantitative form requires the measurement of environmental and physiological processes on the one hand, and psychological events on the other.

The sciences of physics and physiology have already worked out the methods and units which we need for the measurement of stimuli and organic processes, and the psychologist encounters little difficulty there. If, for example, he is working with the law mentioned above, relating the intensity of light to the intensity of the consequent visual experience, he can determine light-values in the manner in which they are measured in

---

This chapter was written by Dwight W. Chapman of Harvard University.

physical laboratories. The units, however, will not serve for measuring the intensity of the experience. The psychologist has been compelled, therefore, to develop methods of measurement of his own in order to complete the equations which are his laws.

In consequence, we shall treat in this chapter some of the most generally used methods of measurement which are typical of psychology.

STATISTICAL METHODS AND MEASURES

Let us suppose that we wish to measure the "range of attention" of a certain observer for some material such as letters of the alphabet. One way of doing this is to arrange a device for exposing very briefly a visual field over which a number of printed letters are scattered at random; the observer is instructed to report immediately after the exposure which of the letters he could recognize in so short a time. We may then determine how many of the letters that he reports are correct, in the sense of having been present in the exposed stimulus field; and this number we may consider to be his range of attention for material of that type.

Strictly speaking, however, if the observer should report five letters correctly after the first exposure, five would be his range for that exposure only. Whether he would invariably report five elements correctly would remain to be investigated and would require us to expose a considerable series of such stimuli. If we thus extend the experiment, we find that his range of attention varies considerably from exposure to exposure, so that in a series of twenty observations we may obtain successive ranges like these:

5, 6, 5, 5, 3, 6, 5, 4, 7, 4, 5, 4, 3, 6, 4, 5, 6, 4, 4, 5

Now it is apparent that any one of these ranges is scientifically uninteresting in itself: it forms only one datum in a collection which we wish to contribute to our knowledge

about a *generalized* range of attention. How, though, are we to make this inductive step?

It is precisely at such a point that all sciences have recourse to statistical devices. The science of statistics is a body of method and fact devised for the purpose of dealing with collections of numerical data—for organizing them into meaningful form and for accomplishing deductions from them. Some of the statistical methods are so common and so obvious that their use would occur to anyone likely to deal with problems of measurement, although he might completely fail to appreciate the logic and assumptions involved in them. Other methods are less well known, and some are extremely difficult. But we may illustrate the meaning of statistical procedure by returning to the above experimental data and treating them in two patently sensible ways.

In the first place, these data would be easier to comprehend if, instead of being strung out in a haphazard series, they were neatly classified. We therefore rearrange them, by writing down in a column the various values that occur in the series, and then writing down opposite each value (or *score*, technically) the number of times it occurs in the series as a whole (again in the language of statistics, its *frequency*). If then we use the customary column-headings, X for a score and f for its frequency, we have Table I.

TABLE I

X	f
3	2
4	6
5	7
6	4
7	1
	$N = \Sigma f = 20$

Such a table is known as a frequency distribution. In constructing this distribution, we have performed one of the elementary processes in statistics—the gathering of data into a

form that is neatly represented and easily understood. A further task of statistics is to describe the properties of such a distribution, as for example its *population*, which is defined as the total number of items which it contains, and is represented by the symbol N. Inspection of the table will show that N can be obtained by adding up the frequencies in the second column. Its value is entered at the bottom of the table, together with its simple formula,

$$N = \Sigma f.$$ [1]

We wish now to pass from a consideration of the various ranges of attention exhibited in this distribution to the notion of a single range of attention. If we hold that such a single, generalized range exists, then it is apparent that each of our twenty obtained ranges measures that true range with some error. At the moment of each exposure of the stimulus, we may suppose, a variety of conditions which varied at random and which we could not control, contributed to an apparent range of attention which was in some cases less than the true range, and in some cases greater. What, then, is the probable magnitude of the true range?

It would probably occur to almost anyone to take the *average* of these obtained ranges as the best approximation to the unknown true range. And that is exactly what a statistician would advise; for he can demonstrate from certain postulates that the average of a series of measures, all of which are suspected of being influenced by random errors, is the likeliest value of the true phenomenon measured. The branch of statistics which this rule illustrates is known as inductive statistics, in contrast to so much of the method as is concerned with the description and classification of data. Inductive statistics takes as its task the inferring of true values

---

[1] The symbol Σ (capital *sigma*) denotes the sum of all quantities in the class denoted by the symbol which follows Σ. Thus, for instance,

$$\Sigma f = f_1 + f_2 + \ldots + f_n,$$
$$\Sigma fx = f_1x_1 + f_2x_2 + \ldots + f_nx_n,$$
$$\Sigma x^2 = x_1^2 + x_2^2 + \ldots + x_n^2.$$

from series of measurements which are known to be incomplete or uncontrollably erroneous.

To return to our concrete example, let us then calculate the average of the obtained ranges in Table I. The method of obtaining an average (or *mean*, as it is better called) is to add together all the items and divide by their number (the population). This can be most easily performed by multiplying each score by its frequency, adding together these products and dividing their sum by N. The formula for the mean (M) is therefore

$$M = \frac{\Sigma fX}{N}.$$

Carrying out this calculation, we have Table II.

TABLE II

X	f	fX
3	2	6
4	6	24
5	7	35
6	4	24
7	1	7
		$\Sigma fX = 96$

$$M = \frac{\Sigma fX}{N} = \frac{96}{20} = 4.8$$

The range of attention for this observer is therefore, as nearly as we can calculate it, 4.8. It must be understood that this figure is an approximation to the ideal mean of an infinite series of measurements, and its accuracy of approximation would increase steadily with an increase in the number of measurements made (*i.e.*, with an increase in N). But it is also perfectly proper to say that 4.8 is the best approximation to the ideal true value that can be made from these twenty data, in the sense that it is more likely to equal the true value than is any other guessed value. Therein specifically lies the function of

statistics—to make plausible guesses from mere samples of data with as little error as the paucity of the evidence permits.

A little reflection will show that the effort to measure any true value in any science must necessarily involve statistical procedure, and that consequently a reliance upon statistics is not unique to psychology. But the reason that statistical method deserves more attention in a discussion of psychological measurement is that different sciences vary in the closeness of approximation with which they are able to measure. In certain branches of physics, and especially for certain purposes, the errors of measurement are so slight as to be practically negligible. In astronomy, on the other hand, where it is necessary to triangulate the universe from the relatively short base line of the diameter of the earth's orbit, reliable results are reached only after a multitude of divergent observations—and these results are reliable only within certain probable limits. This situation is ever so much more notable in psychological measurement, where the extreme complexity of living organisms contributes incessantly to the variability of measurements. For this reason, psychologists have to be everlastingly conscious of the statistical nature of their results and everlastingly cautious of exceeding the limits of predictability indicated in their methods.

The normal distribution. Instead of presenting the data in Table I by means of columns of figures, we might have put it into graphical form. In order to do this, we draw a horizontal base line (the *abscissa*) along which at equal intervals we mark off successive X values. Over these X values we then erect rectangles, each with a unit base and altitude equal to the frequency of the corresponding value of X. This gives us what is known as a *histogram*, as shown in Fig. 9.

If now we connect the center of the top of each rectangle with the next by a continuous line, we have a frequency curve. Such a curve may be of almost any shape, but it is an interesting fact that many natural phenomena, when ordered into a frequency distribution according to some measurable characteristic, give a frequency curve that is high in the center

and that falls off on either side of the center in a shape much like the silhouette of a bell. The ideal form of this curve is known as the *normal* distribution curve. It is shown in Fig. 10.

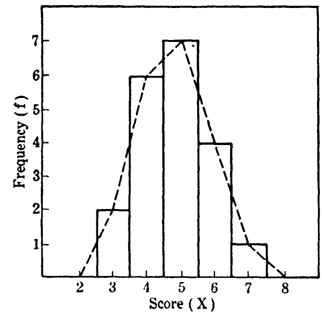


Fig. 9. Histogram (Solid Line) and Frequency Curve (Dotted Line) Representing the Distribution of Scores in Table I.

The normal distribution curve, although it is symmetrical about its midpoint (the mean), may be high and narrow, or low and wide, according as the scores of which it is composed tend to cluster closely about the mean or to lie more widely scattered. This clustering or scattering (dispersion) is a property of distribu-

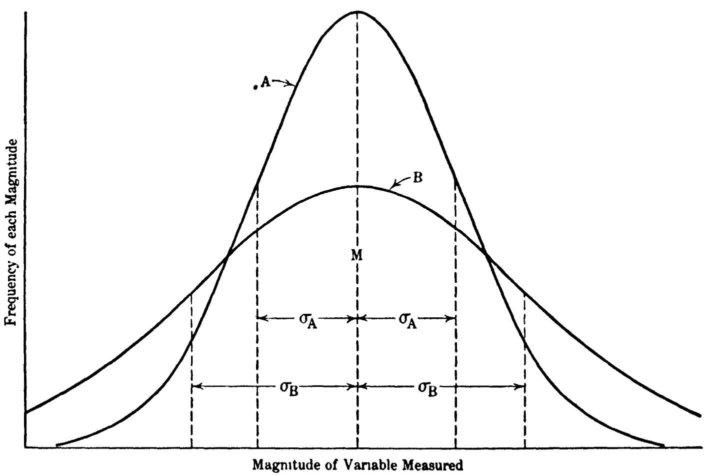


Fig. 10. Two Normal Distribution Curves (A and B) with Standard Deviations in the Ratio 3:5.

Since the area under the curve between any two verticals represents the proportion of variables having magnitudes between those limits, it is easily seen that σ is a sensitive measure of dispersion.

tions which is independent of the value of the mean, and we must therefore note how it is measured. The most common

measure of the scatter of scores about their mean is the *standard deviation,* symbolized by σ (Greek *sigma*). The standard deviation of a distribution is obtained by rescoring each item as a deviation from the mean, squaring each deviation, adding together these squares, dividing their product by N and taking the square root of the quotient. So that, if we let x stand for the difference between any score and the mean (that is to say, $x = X - M$), we have the formula for the standard deviation:

$$\sigma = \sqrt{\frac{\Sigma f x^2}{N}}.$$

Table III shows the actual calculation of σ by this formula, as applied to the distribution in Table I.

TABLE III

X	f	x	x^2	fx^2
3	2	−1.8	3.24	6.48
4	6	−0.8	0.64	3.84
5	7	0.2	0.04	.28
6	4	1.2	1.44	5.76
7	1	2.2	4.84	4.84

$M = 4.8.$ $\Sigma f x^2 = 21.20$

$$\sigma = \sqrt{\frac{\Sigma f x^2}{N}} = \sqrt{\frac{21.20}{20}} = \sqrt{1.06} = 1.03.$$

The value of σ for any distribution shows, therefore, in what way the scores tend to fall around their mean. If σ is large, it indicates a wide spread of items from the mean outward; if σ is small, the tendency is toward a close clustering about M. Of the practical usefulness of σ, two cases may be noted. The first is that, obviously, a prediction from a mean, the σ of whose distribution is large, is less accurate than from a mean about which the individual measures are not so dispersed. Formulas are known which give the probable range within which a predicted mean is accurate enough to be relied upon; and these formulas involve the standard devia-

tion of the distribution from which the mean is taken. A second important use of the standard deviation is in the calculation of coefficients of correlation, which we next describe.

Correlation. A problem beyond the calculation of the central tendency and the dispersion of a distribution is the one of determining whether two distributions of paired scores are correlated. By correlation we understand mutual dependence. For example, if we have measures of height and weight for a group of individuals, we should in general expect that an individual with a height greater than the average of the group would also have a weight greater than the average. If this co-variation occurred, we should speak of height and weight as being directly correlated. On the other hand, if through some strange circumstance the tallest individuals tended to be the lightest, and the shortest individuals the heaviest, we should again regard height and weight as correlated, but we should speak of the correlation as inverse (or negative). A third case would be that in which we found no tendency for a given height to be associated with any particular weight, so that no prediction would be possible from one to the other. This case we should speak of as an example of no correlation, either direct or inverse.

Besides directions of correlation, there are also degrees. If every man 5.5 ft. tall weighed exactly 135 lb., every man 5.6 ft. tall weighed exactly 140 lb., and so on, so that given a man's height we could exactly predict his weight, height and weight would be perfectly correlated. But most correlations fall considerably short of perfection. Height is correlated with weight obviously, but only roughly: it is possible but not probable, that a very short man will weigh a great amount. Thus we can see that there are all possible degrees of correlation, from none at all to perfect co-variance.

It would be convenient, therefore, to be able to rate correlations on a scale running from negative unity through zero to positive unity. Negative unity would then indicate perfect inverse correlation, negative decimals would indicate less per-

fect negative correlation, zero would indicate utter absence of correlation and positive decimals would show direct degrees of correlation up to perfect direct correlation at positive unity. And this is what we are enabled to do by means of the correlation coefficient.

The correlation coefficient. The most significant measure of correlation is the product-moment coefficient, for which the symbol r is used. Although a strict derivation of r is beyond the compass of this chapter, its form and logic can be shown by the following reasoning.

Let us consider two series of numbers (X and Y) paired in three different ways as in Table IV.

TABLE IV

(1)		(2)		(3)	
X	Y	X	Y	X	Y
1	10	1	11	1	12
2	11	2	10	2	11
3	12	3	12	3	10

In the first pair of series, the X and Y values are perfectly positively correlated, since they progress in the same order of magnitude, and each Y value can be exactly predicted from its corresponding X value by the simple relationship, $Y = X + 9$.

The second pair are again positively correlated, but to a less degree. The third pair are perfectly inversely correlated, the two series running in opposite directions of magnitude but in perfectly regular fashion: $Y = 13 - X$.

Now let us consider the X series of the first pair as a distribution having a mean of 2, and let us rescore each of its items as x, the deviation from that mean. Let us then do the same for the first Y series, whose mean is 11, and let us rescore it in terms of y deviations. Now let us multiply together the members of each pair of x and y values, and finally add together all these products, being careful to regard their algebraic signs. This sum we may designate as Σxy. And if

we investigate Σxy, for all three pairs of series, we obtain the results in Table V.

TABLE V

(1)			(2)			(3)		
x	y	xy	x	y	xy	x	y	xy
-1	-1	1	-1	0	0	-1	1	-1
0	0	0	0	-1	0	0	0	0
1	1	1	1	1	1	1	-1	-1
	$\Sigma xy = 2$			$\Sigma xy = 1$			$\Sigma xy = -2$	

An inspection of Table V shows that the quantity Σxy is:

(*a*) Maximal and positive for the case (1) of perfect direct correlation.

(*b*) Maximal and negative for the case (3) of perfect inverse correlation.

(*c*) Intermediate and positive for the case (2) of imperfect direct correlation.

The student, by substituting other series of numbers for those used in this illustration, can easily demonstrate that the value of Σxy will be positive for all cases of direct correlation and negative for all cases of inverse correlation. But he will also find that Σxy will attain values much greater than ± 2 if the x and y series have either more members (a higher N), or a greater spread about their means (a greater σ). This effect is irrelevant to the measure of a correlation, and correction must be made for it. This can be done quite simply by dividing Σxy by the N common to the paired series and by the standard deviations of both series (σ_x and σ_y). This gives us at last the formula for the coefficient of correlation:

$$r = \frac{\Sigma xy}{N\sigma_x\sigma_y}.$$

This coefficient has the following very convenient properties which we have said would be desirable:

1. It always lies between the limits $+1$ and -1.

2. Its absolute value shows the degree of correlation, from no correlation whatsoever at zero to perfect correlation at 1.00.
3. It is negative for all degrees of inverse correlation, and positive for all degrees of direct correlation.

We can illustrate the calculation of a correlation by a hypothetical example. Suppose that we have two tests, the first a test of ability in use of English, and the second a test of ability in foreign languages. We wish to know whether these two abilities appear to be connected—that is to say, whether performance on the one test correlates with performance on the other. Our method would be to give the two tests to the same group of pupils and record for each pupil his mark on each test.

An imaginary result is given in Table VI, showing the scores of ten pupils on the two tests. The calculation of the means and standard deviations of the two series of scores is not shown, since they have already been illustrated. The com-

TABLE VI

Pupil	X (Score on first test)	Y (Score on second test)	x	y	xy
I	1	2	−5	−4 5	21.5
II	2	5	−4	−1 5	6 0
III	4	7	−2	0 5	−1 0
IV	4	4	−2	−2 5	5.0
V	7	6	1	−0 5	−0.5
VI	8	9	2	2 5	5 0
VII	8	6	2	−0.5	−1.0
VIII	8	9	2	2.5	5.5
IX	9	9	3	2.5	7.5
X	9	8	3	1.5	4.5

$M_x = 6 \quad M_y = 6\ 5 \quad \sigma_x = 2.83 \quad \sigma_y = 2.25 \quad \Sigma xy = 52.0$

$$r = \frac{\Sigma xy}{N\sigma_x\sigma_y} = \frac{52}{10 \times 2.83 \times 2.25} = 0.82.$$

putation of the quantity Σxy is shown in detail, as well as the final calculation of r, the coefficient of correlation.[2]

Since the correlation, when calculated, turns out to be positive and well up toward 1.00, we should conclude that performance on these two tests was highly related in such a way that an individual who scored above the average on the one test would be very likely to do better than the average on the other.

One scientific value of determining correlations results from the fact that, when two phenomena are found to co-vary consistently, the hypothesis becomes plausible that the two phenomena are connected by the possession of some common factor. Thus, for example, if tests of motor skill, ability in arithmetic, understanding of verbal relations and memory span all are found to be positively correlated with one another, the indication is that there probably exists some ability or abilities more general than these specific ones and psychologically basic to them. It is precisely on the ground of such findings that some psychologists argue for the recognition of general abilities which affect specific mental performances of all sorts.

The method of correlation is also used to perfect and refine tests themselves. The intelligence test which was devised during the World War was found to measure intelligence satisfactorily for the purposes for which it was intended, provided those who took the test were literate and spoke English. But the intelligence of those who had not this fundamental training was obscured by the effect of linguistic difficulties. It became desirable, therefore, to construct a test which should measure intelligence largely through the medium of motor acts of a non-verbal sort. Such a test was actually devised; but, of course, before it was put into use, it had to be examined for consistency with the original test. The method of

[2] Although, for the sake of simplicity of exposition, our imaginary example shows test scores from ten individuals, the correlation coefficient r would almost never be calculated from so few data. In any actual study one would hardly rely on a coefficient calculated from less than fifty to a hundred scores.

thus testing a test for consistency is quite clear; the new test and the old standard test are administered to the same large group of people. The scores on the two tests are then made the data for a coefficient of correlation. If a high positive correlation is obtained, the second test obviously measures with little error what the first test measured, and it is in that sense reliable. If the correlation is low, the second test must be analyzed and reconstructed to bring it into agreement with the first.

PSYCHOPHYSICAL MEASUREMENT

Having outlined the most important general statistical procedures in psychological measurement, we must turn now to the metrical problem which is most peculiar to psychology—that of psychophysical measurement. We have already said that the psychologist's special province is the determining of relationships between experience and the environment, and we have pointed out that the ability to measure experience is essential to a discovery of the exact form of these relationships. But the measurement of experience presents important difficulties, from which have arisen special metrical procedures.

It is apparent, first of all, that experience varies in respect of what we may call certain dimensions. Visual experience, for instance, shows a brightness variation from very dark to very bright, which immediately suggests the conceptual existence of a brightness continuum at some position along which any given element of the visual field could be placed. Again, tones have an insistent pitch quality (highness-lowness), and the circumstance of a musical scale shows us the practicability of thinking of any one tone as having position in a pitch continuum. Indeed most, if not all, of the aspects of experience offer themselves as dimensional affairs, along which particular impressions can be ordered according to their degree.

Now in physical measurement we deal with quite analogous continua—those of length, mass and time; but the selection of physical units wherewith to measure distances along the

continua (centimeters, grams and seconds) is vastly easier. We need only decide on some convenient object which has an invariable position in one of the dimensions (such as the distance between two scratches on a platinum bar in Paris, which is *the* meter) and then by manipulating this object find out how many times our standard is contained between the zero point of its continuum and some point which we wish to measure.

The analogous procedure in measuring sensory experience would seem to be to select some impression as a standard and to discover how many times it is contained in a second impression. However, two great difficulties prevent us here. First of all, whereas a standard physical measure is selected for its permanence and relative invariability, impressions are evanescent. If we were to select today a given impression of light in a given observer as the unit of brightness, there is no way in which this impression could be earmarked so that the observer might recognize and use it if it recurred tomorrow. And secondly, it is a notable property of such experiences that, although they can be regarded as having relative positions on a conceptual dimension, those which are high in position do not seem to be in any real sense multiples of those which are lower. An experienced bright light is easily judged as more intense than a weak one; but the judgment that it is so-and-so many times as bright is difficult to make with assurance. As a consequence, the common metrical method in the physical sciences is ill adapted to psychological phenomena.

The differential threshold. The solution of this problem has been to measure distances on the experiential continua not in terms of multiples of some standard impression, but in terms of a standard difference between impressions. We have said that the judgment that two impressions are different in respect to some dimension is a feasible task to set an observer. It needs now to be added that, if we take two different stimuli and let them approach each other physically, the observer will cease to experience a difference between their impressions some

time before the stimuli have actually achieved physical equality. In other words, introspective discrimination is considerably less accurate than physical measurement.

In technical terms, the difference between two stimuli necessary to produce two discriminable impressions is known as the *differential threshold* or *limen* (abbreviated, D.L.) for the dimension along which the discrimination is to be made (hue, brightness, pitch, weight, etc.). And this smallest possible difference between the impressions, which results from a liminal stimulus-difference, is known as a *just noticeable difference* (abbreviated, j.n.d.).

Since the j.n.d. is a difference which can be empirically established by manipulating stimuli, it offers itself as an acceptable unit for the measurement of experience and has been universally adopted for that purpose. To be most useful as a unit, however, the j.n.d. ought to be a constant, like the meter, the gram or the second. Historically, psychophysics has been marked by an interesting controversy as to whether j.n.d.'s at all points of the range of stimuli and impressions are actually invariable; and it has been pointed out that although all j.n.d.'s are alike in being 'just noticeable,' there is no *a priori* reason why just-noticeability might not vary with the magnitude of impression for which it was tested. The general and very practical answer has been that, although it is controversial whether the j.n.d. is in fact constant, or whether its constancy can even be tested, the purposes of measurement are adequately served if the j.n.d. be simply and arbitrarily defined as a constant unit.

THE PSYCHOPHYSICAL METHODS

It must be realized that, although we speak of a differential limen as that difference between two stimuli which will bring about a j.n.d. in experience, the limen is really a statistical concept. Let us suppose that two weights are presented to an observer to lift with the hand, both of them weighing physically 100 gm. Let us now, without the observer's knowledge,

increase one of the weights gradually by small increments, presenting the two weights after each increment and requiring the observer to report whether they seem equal or different. We know from the fact of the limen, alluded to above, that the first few increments will not evoke a judgment of 'different.' But at some point as we increase the one stimulus, say at 107 gm., the observer will cease to judge the two weights as equal, and will begin giving us a series of judgments of difference.

This point, where the judgment changes, we might be tempted to regard as the differential threshold for weight in the region of 100 gm. However, if we hope to confirm this result exactly by a repetition of the experiment, we shall almost certainly be disappointed, since a second series of liftings of the weights may quite easily show the judgments to change not at 107 gm., but at, perhaps, 103 gm., or at 110 gm. Apparently, then, the sensitivity of the observer is subject to small but measurable changes, which occur in the same random manner that characterizes the errors which creep into all measurements. Consequently our determination of a single-valued threshold for weight must be a statistical deduction from a variety of observed values, exactly as was our measurement of the range of attention.

Several ways of collecting and treating the data for the calculation of differential thresholds have come down to us from the earlier psychophysicists and are in continual use. Together they are called the psychophysical methods.

All the psychophysical methods have in common the characteristic of employing a standard stimulus and a variable comparison stimulus which is to be compared by the observer with the standard stimulus. The methods fall into two subgroups, however, according as the comparison stimulus is varied by numerous very small steps over the necessary range (or continuously) or by rather large, constant steps. In the first group belong the *method of limits* and the *method of average error*; in the second group belongs the *method of constant stimuli.*

The method of limits. In this method two stimuli are presented—a variable stimulus and a standard stimulus with which the variable is to be compared. The experimenter may begin with the variable stimulus at a value somewhat smaller than the standard, and then increase the variable by successive steps, calling upon the observer to judge it in relation to the standard after every such increment. The point above the standard at which the variable begins to be judged greater is recorded. Next, the comparison stimulus is made considerably greater than the standard, and the two are again presented over and over to the observer, while the comparison stimulus is gradually reduced. This time the value of the comparison stimulus at which the judgment changes from one of difference to one of equality is recorded. Each of these so-called ascending and descending series is repeated a considerable number of times, since each, as we have noted, is subject to errors and can enter into the final computation only as one of many data.

Having completed, say, fifty of both series, the limen is calculated by taking the mean of the ascending data, the mean of the descending data, and then averaging them.

The method of average error. In this method the observer himself controls the magnitude of the variable stimulus. He begins with a variable which is distinctly either greater or smaller than the standard, and he varies it until he is satisfied with the subjective equality of the two. The difference between the variable stimulus after his adjustment and the standard stimulus is recorded after each adjustment, and these errors are tabulated for a considerable series. At the end of the experiment their mean is calculated. Obviously the average error will be large where discrimination is poor, and vice versa; and thus the average error like the threshold may be taken as a measure of sensitivity.

The method of constant stimuli. Of the various psychophysical procedures, the method of constant stimuli is at the present time the most refined and the most frequently re-

lied upon. The other two methods employed variable comparison stimuli which were increased or decreased by a large number of small steps or continuously. The method of constant stimuli employs a limited set of fixed comparison stimuli set at regular intervals above and below the standard stimulus. These stimuli are never varied during the experimental session, but they are presented in haphazard order to the observer, each following a presentation of the standard; and the observer is usually required to render in each case a judgment *greater, less* or *equal.* From the proportion of judgments of each of these three categories given in response to the various stimuli, the threshold is calculated.

A tangible example will make the procedure clear. Let us suppose that we wish to measure the threshold for lifted weights in the region of 100 gm. We prepare a standard stimulus of 100 gm., and perhaps seven comparison stimuli weighing 88, 92, 96, 100, 104, 108 and 112 gm. We then present the standard stimulus to be lifted, and immediately following it, one of these comparison stimuli; and the observer judges the comparison stimulus to be heavier, lighter or equal to the standard. We note down his judgment opposite the value of the comparison stimulus, and repeat the procedure, selecting another comparison stimulus. The order of selection of the comparison stimuli is kept irregular, so that the observer shall not be influenced by any expectation of regularity; but all the stimuli, by the end of the experiment, will have been presented to him an equal number of times.

Suppose, then, that we have completed 100 presentations of each stimulus, and that our results are as in Table VII. These results are still clearer when presented graphically, as in Fig. 11. Such curves are called *psychometric functions.*

Let us take for analysis the psychometric function representing the frequencies of greater judgments. We see at once that even the comparison stimulus weighing 88 gm. has been judged heavier than the standard of 100 gm. in two of the hundred cases. With successively heavier comparison stimuli, the number of "greater" judgments increases rapidly until, at 112 gm.,

all of the hundred judgments are of this category. Obviously the limen for judgments of greater ought to lie somewhere

TABLE VII

	Weight of comparison stimulus in grams						
	88	92	96	100	104	108	112
Number of times judged greater....	2	9	29	47	78	97	100
Number of times judged equal.... .	8	23	40	38	17	2	0
Number of times judged less......	90	68	31	15	5	1	0
Total judgments	100	100	100	100	100	100	100

between these two points. The most logical place for the limen would be at the value of the comparison stimulus which would as often as not be judged greater than the

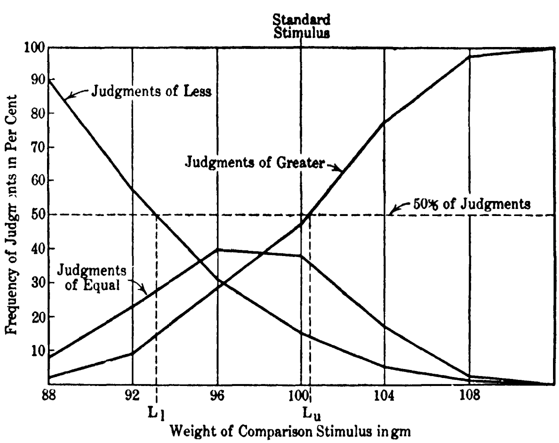


FIG. 11. PSYCHOMETRIC FUNCTIONS. GRAPHICAL REPRESENTATION OF THE DATA OF TABLE VII.

The asymmetrical disposition of the limens (Ll and Lu) about the standard stimulus (100 gm.) results from the 'time error,' which occurs when the standard stimulus precedes the comparison stimuli and which has the effect of making a present comparison stimulus seem more effective in comparison with a past standard stimulus.

standard—that is to say, would be judged greater in exactly 50 out of 100 cases. Therefore, the limen must lie between 100

gm., which was judged greater 47 times, and 104 gm., which was judged greater 78 times. If we now draw the horizontal dotted line which represents 50 of the hundred judgments, we find it intersecting the curve of greater judgments at a point over 100.43 gm., which we may take roughly to be the limen for greater judgments, or upper limen (L_u).

In the same way we can calculate from the curve of less judgments a lower limen (L_l), which indicates the value of the comparison stimulus which will give an impression of weight just noticeably less than that given by the standard stimulus. The method of constant stimuli thus affords a convenient way of determining a limen both above and below a given value of stimulus.

In actual practice, a more refined method of treating the results is used. The assumption is made that the points in our diagram which we have connected by straight line segments could be better fitted by a smooth curve. The mathematics of this method, however, is beyond the scope of this book, and the general procedure is not fundamentally different from the one just outlined.

REFERENCES

On general statistical method:

1. BROWN, W., and THOMSON, G. H. *Essentials of mental measurement.* Cambridge, England: University Press, 1925.
2. GARRETT, H. E. *Statistics in psychology and education.* New York: Longmans, Green and Co., 1926.
3. KELLEY, T. L. *Statistical method.* New York: Macmillan Co., 1924.

On the psychophysical methods:

4. MYERS, C. S., and BARTLETT, F. C. *Textbook of experimental psychology.* Third edition. New York: Longmans, Green and Co., 1928.
5. TITCHENER, E. B. *Experimental psychology.* New York: Macmillan Co., 1905. Vol. II, part I.

See also Brown and Thomson (1), *supra.*

CHAPTER 4

VISION

With this chapter we begin a survey of the world of sensory experience—of the world as it is given to us through the channels of the various senses. From the point of view of physical science, the world is a complex pattern of physical energies and forces; but to the experiencing person himself, it is made up of such psychological realities as colors, tones and odors. As students of psychology, we are interested in the analysis and description of this sensory world, and in the laws which determine its structure and properties.

Sensory phenomena must always be clearly distinguished from the physical agents, or stimuli, which give rise to them. Thus, a color is something quite different from a light-wave. Nevertheless, sensory phenomena are related to their stimuli by definite laws of dependence. These laws, in so far as they have been worked out, form the subject-matter of the next few chapters.

In the study of any one of the special senses, it is convenient to begin with the problems of *quality* and *intensity*. Tones may be deep or shrill, tastes may be sweet or saline; differences of this sort are differences in regard to quality. But tones are also faint or loud, and tastes are also weak or strong; here we have to do with differences of intensity.

Our study of sensory experience begins with the domain of vision. In this chapter, we shall be concerned with problems of visual quality and intensity.

In the field of vision, the demarcation between quality and intensity is a less simple affair than in the other sensory

---

This chapter was written by D. M. Purdy of the University of Maine.

fields, but this difficulty need not concern us in the present chapter.

COLORS AND THEIR PROPERTIES

Throughout this chapter we shall be occupied with the study of colors. For us the word *color* always means a particular sort of sensory experience, never the physical light-waves which are responsible for this experience. We shall have occasion, eventually, to inquire into the relationships between colors and their stimuli, but at the outset we must turn our attention to the description of colors as such—to the various ways in which one color can resemble or differ from another color. Such a description is purely psychological and essentially independent of any knowledge whatever about light-waves.

We use the word color in a broad sense, so that it includes both the *achromatic* and the *chromatic* colors. The achromatic group comprises white, the various shades of gray and black. The chromatic group includes such colors as red, yellow, green and blue.

Besides being classified as achromatic or chromatic, colors can also be grouped together according to their so-called *modes of appearance.* Most of the colors which are important for us in daily life are *surface-colors.* For example, a sheet of paper or a piece of cloth, when viewed under ordinary lighting, displays a surface-color. A color of this type has a compact appearance; it seems to *belong to* an object and to have a very definite spatial location on the surface of that object. (See pp. 284-286 on the mode of appearance of color in relation to perceiving in general.)

On the other hand, if we look at a small spot of light in a room which is otherwise completely dark, we experience a *film-color.* Film-colors have a less compact appearance than surface-colors. Their apparent location in space is more or less vague and indefinite. They impress one as isolated color phenomena rather than as characteristics of a particular object. The blue of the sky is an example of a film-color.

From many points of view the film-colors are simpler phenomena than the surface-colors; and we shall proceed first to consider the various properties of film-colors. Where the word *color* appears in this discussion, *film-color* is implied. The surface-colors will receive separate treatment later.

Let us begin with the achromatic film-colors. We can think of all possible colors of this type as arranged in order in a series, with the most brilliant white at one extreme, the deepest black at the other extreme and the light and dark grays between them. In such a series the colors are said to be arranged in order of their *brightness* (or brilliance).

Chromatic colors also have the attribute of brightness. Imagine comparing a chromatic color, let us say some particular red, with the whole series of achromatic colors from extreme white to extreme black. We can at once place the red color somewhere on this brightness scale. We observe, for instance, that it is not so bright as the brightest, and not so dark as the darkest, of the achromatic colors; and we can go further and find an achromatic color of some particular brightness which our sample of red resembles more closely than any other in the series. By the brightness of a chromatic color we mean that attribute in respect of which it may be matched, or equated, to some particular achromatic color.

Chromatic colors not only possess brightness, but they also possess two other attributes which achromatic colors lack, namely, *hue* and *saturation*. When we describe a color as red, or as bluish green, for example, we are naming its hue. Two reds may be exactly alike both in hue and in brightness, and yet differ in richness or saturation. The one, for instance, may be a rich, full red, while the other has only a mere tinge of reddishness and is barely distinguishable from an achromatic color. The difference between reds of high and low saturation is so pronounced that we often use a special word *pink* to describe certain weakly saturated reds. The saturation of any chromatic color may be defined as its degree of difference from an achromatic color of the same brightness. We can think of an achromatic color as a color of zero saturation; it is a limit-

ing point which chromatic colors approach as they become less and less saturated.

We turn now to the attribute of hue. As a convenient starting point, let us consider the appearance of the *spectrum*, the band of colors which is seen when a beam of sunlight is split up into its component wave-lengths by means of a prism. In the spectrum we have a series of hues arranged in orderly sequence, with red at the left or long-wave end, and violet at the right or short-wave end.

We may regard this series as consisting of four principal sections. We begin, at the extreme left, with a red. After it comes a sequence of hues which are progressively more yellowish and less reddish, until eventually a yellow is reached which has no trace of reddishness. In the second section we have a series of hues which are intermediate between yellow and green: the greenishness continually increases, and the yellowishness decreases, and finally we have a green which is without a trace of yellow. In the third section, there is still a new direction of change—from green through increasingly bluish greens to blue. The fourth and last section of the spectrum contains hues which are intermediate between blue and red—the bluish and reddish violets.

The spectrum does not contain all possible hues. There are hues intermediate between blue and red which are still more reddish than the extreme spectral violet. These are the purples and carmines. Such hues are not produced by any single wave-length, but require a mixture of wave-lengths; they can be obtained, *e.g.,* by mixing lights from opposite ends of the spectrum. If we add these colors to the fourth section of the spectrum, we have a complete series of intermediate hues which begins with blue and ends with red—our starting point at the left-hand end of the spectrum.[1]

The total series of hues is, therefore, a *closed* series, and can be symbolized by a closed geometrical figure, such as a circle

[1] The extreme red of the spectrum is, to be strictly accurate, a slightly yellowish red. The "pure" or "primary" red, like the purples and carmines, lies outside the spectrum.

(Fig. 12, *a*). In this series there are four hues which have a peculiar significance. Consider, for instance, the hues in the first section of the spectrum. Throughout this section, the hues become steadily more yellowish and less reddish. But when yellow is reached, a new change in hue begins—a change away from yellow and towards green. At the turning point stands a yellow which is neither reddish nor greenish—a *primary* yellow. At the end of the second section is another turning point—a primary green which is neither yellowish nor

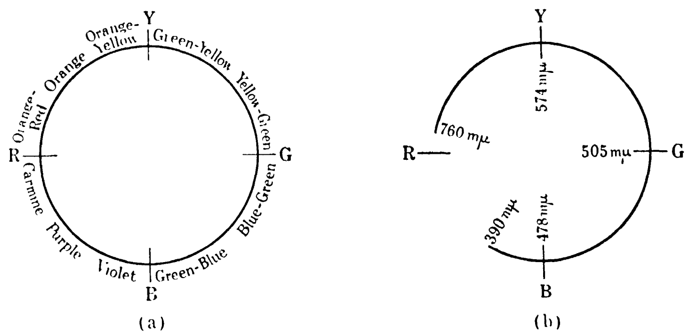


FIG. 12. PSYCHOLOGICAL COLOR CIRCLE.
(*a*) R, Y, G and B represent the four primary hues
(*b*) Relationship of the series of spectral colors to the psychological color circle. The wave-lengths of the spectral limits and of the three spectral primary hues are indicated in the diagram.

bluish. The total hue series contains two other turning points—a primary blue, which lacks greenishness or reddishness, and a primary red which lacks bluishness or yellowishness. These four primary hues[2] are represented by R, Y, G and B in Fig. 12.

In order to do justice to these 'turning points,' a square rather than a circle is sometimes used to represent the hue series, with the four primary hues placed at the corners.

Suppose we wish to construct a geometrical figure which represents all three attributes of color—brightness and saturation as well as hue—and which includes all possible colors,

[2] They are often called the "psychological primaries," since the term "primary hue" is sometimes used in other senses.

both chromatic and achromatic. For this purpose, a solid cylinder may be used (Fig. 13). In this representation, the series of achromatic colors is symbolized by a line, the axis of the cylinder. Extreme white stands at the upper end of the line, and extreme black at the lower. A chromatic color is represented by a point outside the axis. The greater the distance of this point from the axis, the greater the saturation of the color. The height of the point, measured perpendicularly to the base of the cylinder, represents the brightness of the color. Hue varies circumferentially around the axis.

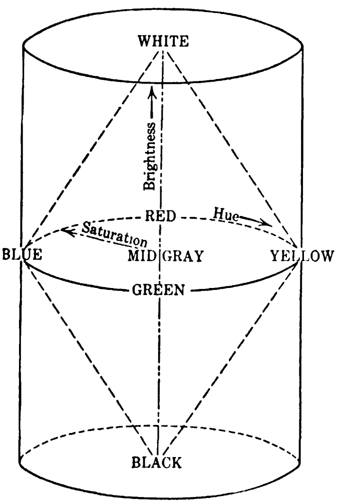


FIG. 13. PSYCHOLOGICAL COLOR SOLID. The solid is drawn sometimes as a cylinder and sometimes as a double cone. After Troland (13).

A line drawn parallel to the axis at any place in the cylinder represents what we shall call a simple brightness series —a series of colors which have the same hue and saturation, differing only in brightness. A line beginning at the axis and drawn perpendicularly to it, at any place, stands for a series of colors which have the same brightness and hue, but differ in saturation—a simple saturation series. A circle of any size, drawn concentric with the axis, indicates a simple hue series, constant in brightness and saturation.

Now, it so happens that very bright and very dark colors are always low in saturation. The richest colors are always colors of intermediate brightness. If we wish to take account of this fact, in an approximate way, we can substitute for the cylinder a solid figure which tapers toward either end, such as the *double cone* shown in Fig. 13. All colors are to be thought of as lying within the boundaries of this cone.

Our discussion thus far applies to the *film-colors.* The properties of film-colors and *surface-colors* have certain important differences which we can now consider.

A sheet of gray paper, viewed in daylight or artificial light under the usual conditions of everyday experience, furnishes a typical example of an *achromatic surface-color.* Our perception of such a color is a more complex affair than it might seem at first thought. We see that the paper is gray; if a white or black paper were substituted for it we should notice the difference. But at the same time, we also perceive the paper as lying in a certain particular sort of *illumination.* If the daylight or lamplight were strengthened or weakened, we should perceive this change *as* a change in the illumination. The paper itself would still seem to have approximately the same shade of gray as before, but it would appear to us as a more strongly or weakly illuminated gray.

Within certain limits, the whiteness, grayness or blackness of a surface-color is experienced as a property which remains *constant* despite changes in illumination. The explanation of this fact is a difficult problem with which we are not especially concerned here. It is usually said that the constancy must depend in some way upon our ability to perceive the color of the surface itself and the illumination on the surface as two distinct aspects of the total situation.

It should be particularly noted that the perception of these two separate aspects, surface-color and illumination, occurs only under certain conditions. Suppose that upon a piece of colored paper in general illumination we project over a part a special illumination—a light or a shadow. Suppose then that we view this specially illuminated part of the paper through a very small hole in an opaque screen, which obscures the rest of the paper. When we introduce the screen, two changes happen: (1) the paper no longer appears as a surface-color, but rather as a *film-color*; and (2) we no longer perceive the illumination and the color of the paper itself as two distinct factors. Thus, the separation of these two factors does not occur when

the visual field is narrowed down to a single patch of color, but only in a more complex situation. When we remove the screen and restore this complex situation, the film-color with its peculiar degree of brightness gives place to a surface-color with two different *kinds* of brightness—the brightness of surface and the brightness of the superposed illumination. (On this constancy phenomenon, see pp. 278-283.)

Similar phenomena occur in the case of chromatic surface-colors and chromatic illuminations. When we look at a piece of red paper in daylight, we see the redness as a property of the surface; the illumination is perceived as achromatic. On the other hand, when we view a piece of white paper in the red light of a photographic dark room, we experience the redness as a property of the illumination, and the paper itself still looks more or less white. And even in the case of a chromatic (*e.g.*, red) object which is illuminated by chromatic light of dissimilar hue (*e.g.*, yellow), we can distinguish, to a certain extent, between hue of surface and hue of illumination.

In describing a surface-color, therefore, we have to introduce a complication which is unnecessary with film-colors. We must differentiate between the hue, brightness and saturation which are experienced as belonging to the surface itself, and the hue, brightness and saturation of its apparent illumination. A piece of white paper in dim light has high surface brightness, but low illumination brightness; and a piece of black paper in strong light has low surface brightness, but high illumination brightness.

STRUCTURE OF THE EYE

The eye (Fig. 14), like a photographic camera, consists of two main parts: a sensitive film at the back, and an optical mechanism which throws an image on this film. Light entering the eye passes first through the transparent *cornea*, next through a watery fluid, the *aqueous humor*, then through the *lens* and finally through a gelatinous substance, the *vitreous humor*, before reaching the sensitive film. The cornea, the lens

and the two humors function together in the formation of the optical image.

Just in front of the lens is the *iris*, a diaphragm containing a circular aperture which automatically changes in size when the illumination changes. This aperture (the *pupil*) thus regulates the strength of light in the optical image.

The lens can be focused for near or far objects. The focusing (or *accommodation*) of the eye is accomplished by an auto-

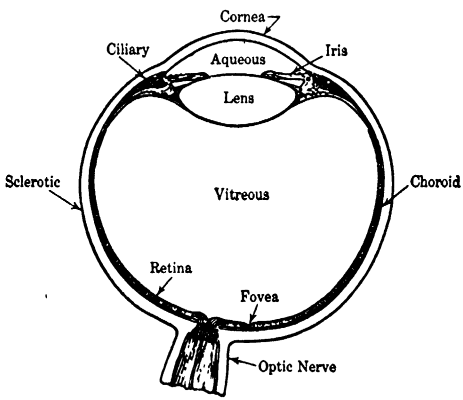


FIG. 14. CROSS-SECTION OF THE HUMAN EYE.
This figure represents a cross-section of the right eye, as viewed from above. After Troland (13).

matic change in the curvature of the lens. The lens has elasticity, which tends to make it bulge; but it is held under tension by a ring of radial fibers encircling it. The tension of these fibers is in turn controlled by an outlying ring of muscular tissue, the *ciliary muscle*. When we look at a nearby object, the ciliary muscle contracts, the tension on the lens is removed and the lens assumes greater curvature. When we transfer our gaze to a distant object, the ciliary muscle relaxes, and the lens is now subjected to a tension which causes its front surface to flatten.

The whole eyeball is enclosed in a leathery casing, the *sclerotic coat*. At the front of the eye this coat has a transpar-

ent portion, the cornea. The sclerotic coat is lined by a dark-colored membrane, the *choroid coat.* Within the choroid, in turn, is a thin membrane which is the sensitive film of the eye and is called the *retina.*

The retina (Fig. 15) has a very complex microscopic structure. It contains two sorts of nerve cells or neurons: the

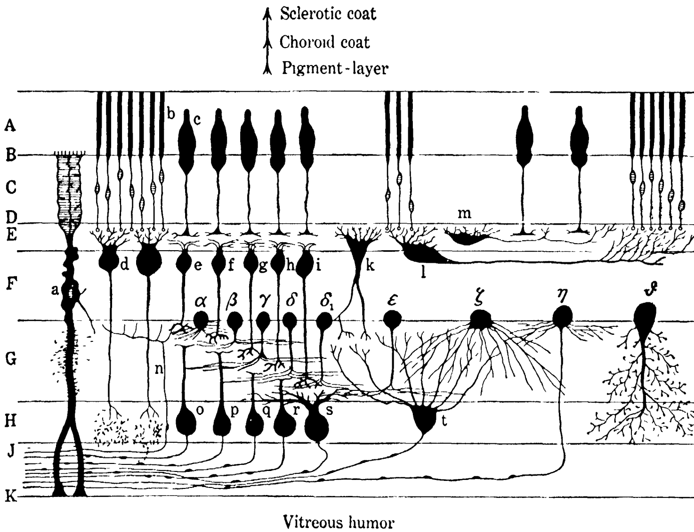


FIG. 15. SCHEMATIC CROSS-SECTION OF THE HUMAN RETINA.
b = a rod.
c = a cone.
d, *e*, *f*, *g*, *h*, *i* = bipolar cells.
o, *p*, *q*, *r*, *s*, *t* = optic-nerve cells.
k, *l*, *m* = lateral neurons (horizontal cells).
α, β, γ, δ, ε, ζ, η, θ = lateral neurons (amacrine cells).
After Kallius; based on the work of Ramón y Cajal.

visual receptors, which are its light-sensitive elements, and a system of neurons for the transmission of nervous impulses from them.

The receptors are of two types, the *rods* and *cones*, so named after their typical shapes. Curiously enough, the rods and cones are situated behind the layers of conducting neurons, so that

the light must first traverse these layers (which are almost perfectly transparent) before it can act upon the receptors.

The rods and cones are connected, by way of synapses, to short conducting neurons which are called *bipolar cells.* The bipolar cells in turn make synaptic connection with a second set of conducting neurons, the *cells of the optic nerve.* The cells of the optic nerve have very long axons, which constitute the individual fibers of the *optic nerve* connecting the retina with the brain.

Whereas there are more than a hundred million receptors in the retina, the optic nerve contains only about half a million fibers. As a rule, a large number of receptors are connected with the same bipolar cell, and, in turn, a large number of bipolar cells share the same neuron in the optic nerve.

Near the center of the retina is a small pit or depression called the *fovea.* The receptors at this place are very small in diameter and are packed very closely together. Each receptor in the fovea has its individual bipolar cell and optic-nerve cell, which it shares with no other receptor. Hence this central region furnishes more distinct vision than the outlying or *peripheral* retina. Moreover, the eye is an imperfect optical instrument, and the images formed on the peripheral retina are less sharp than those formed on the central area.

When an object in the visual field attracts our interest, we involuntarily turn our eyes so that the retinal image in each eye falls on the fovea and the object is seen with maximum distinctness. Such an adjustment of the eyes is called *fixation.*

The rods and cones have a characteristic distribution on the retina. The fovea and its immediate surroundings contain only cones; they make up the central *rod-free area* of the retina. Just outside this area, a few rods are intermingled with the cones, and the rods become more and more numerous as the edges of the retina are approached. The extreme periphery contains an overwhelmingly large percentage of rods, but is not quite free of cones. As we shall see, the rods differ from the cones not only in structure but also in function.

The fibers of the optic nerve pass from all parts of the

retina toward a common meeting place, where they form themselves into a bundle, the optic nerve. The place of exit of the optic nerve is called the *optic disk* or *blind spot*; the optic disk of each eye is situated on the inner or nasal side of the retina. Under ordinary conditions the optic disk is blind, as one would expect if it consisted entirely of nerve tissue and contained no rods or cones. However, recent experiments indicate that the optic disk responds if very strong light is thrown upon it. It seems likely, therefore, that this region contains a few receptor cells.

In addition to the bipolar and optic-nerve cells, which form a direct line of conduction between the receptors and the brain, there are also *lateral neurons* which connect one part of the retina with another (Fig. 15).

Between the receptor layer and the choroid membrane is a layer of dark brown *pigment cells*. These cells have been shown to play a significant part in the process of visual response, although their exact function has not yet been clearly made out.

COLORS AND THEIR STIMULI

The stimulus for vision is an electromagnetic wave-motion of very short wave-length. It is customary to express the wave-lengths of this radiant energy in terms of a unit called the millimicron (abbreviated $m\mu$), which is equal to one-millionth of a millimeter. The longest waves of the visible spectrum, at its extreme red end, are about 760 $m\mu$ in length; the shortest, at the extreme violet end, are about 390 $m\mu$ in length. Under exceptional conditions, *i.e.*, when the stimuli are unusually strong, the range of visible waves extends as far as 800 $m\mu$ and 365 $m\mu$.

There exists a great variety of *invisible* electromagnetic waves which are longer than 800 $m\mu$, for example, the waves of radiant heat and the waves used in radio broadcasting. There are also waves shorter than 365 $m\mu$, such as those of ultraviolet radiant energy and of X-rays. The length of the shortest

waves known to the physicist is about 0.0001 mμ, and that of the longest is about 30 million mμ. All these waves are alike in kind and differ only in length, but the human retina is so constituted that it responds only to a very minute range of these wave-lengths. The property which causes us to give this group of waves the special name of 'light-waves' is a property of the human retina rather than any intrinsic property of the waves themselves.

We describe a given kind of radiant energy as *homogeneous* if it is made up of waves of a single length, and as *complex* (or heterogeneous) if it is composed of a mixture of waves having different lengths. The spectrum is an ordered sequence of homogeneous stimuli of different wave-lengths. The sunlight, or other kind of 'white light,' out of which the prism forms the spectrum, is a complex stimulus. Practically all the visual stimuli which we encounter in common experience are very complex. If, by means of a prism, we analyze the light reflected by some common object, such as a piece of green cloth, or a red flower, we invariably find that the object reflects light of a great number of wave-lengths.

A precise description of a homogeneous stimulus must include two factors: first, the wave-length; and second, the energy value of the stimulus (*i.e.*, the quantity of radiant energy delivered to unit area of the retina in unit time).

A complex stimulus may be considered as made up of a number of stimuli of particular single wave-lengths, each having its own energy value. A complete description of such a stimulus must include the individual wave-lengths which compose it, and the energy value of each component wave-length.

If we confine ourselves to homogeneous stimuli, we have a rather simple relation between *hue*, on the psychological side, and *wave-length*, on the physical side—as the spectrum shows. Fig. 12, *b*, illustrates the fact that the sequence of homogeneous stimuli from 760 to 390 mμ yields a sequence of hues which includes almost, but not quite, a complete hue circle. The figure also indicates the wave-lengths of the three

spectral stimuli which furnish primary hues—the primary yellow, green and blue.

When the visual stimulus is complex rather than homogeneous, it is impossible to formulate any such simple relationship between the hue of the color and the characteristics of the stimulus. We shall consider this problem when we come to deal with the facts of stimulus mixture (pp. 78-83).

Let us next consider the attribute of *saturation*. As a rule, the colors from homogeneous stimuli are more saturated than those from complex stimuli. No complex stimulus furnishes colors more saturated than those of the spectrum; and most complex stimuli give decidedly less saturated colors than those of the spectrum. But here again it is impossible to express the correlation between color and stimulus in simple form; and furthermore, an acquaintance with the facts of stimulus mixture is necessary to an understanding of the problem.

When the *energy* of a light-stimulus is increased, one experiences an increase of *brightness*. We cannot, however, say that brightness depends simply on the amount of energy. This statement is true only so long as it refers to stimuli of the same composition of wave-lengths. As we shall see, brightness depends on wave-length as well as energy. Generally speaking, two stimuli which are equal in energy but different in wave-length produce unequal brightnesses.

In the case of chromatic colors, moreover, a variation in the energy of the light-stimulus carries with it a change not only in brightness but also in *hue* and *saturation*. When the stimulating energy is weak, the resulting colors have low saturation. As the energy is increased, the saturation of the colors increases up to a maximum, but diminishes with further increase in energy (see p. 72). The hue produced by any given kind of stimulus also changes slowly but steadily with change of energy. Reds, oranges and yellowish greens become more yellowish, and bluish greens and violets become more bluish, with increasing strength of stimulus. Thus the hues of the spectrum are not strictly dependent upon wave-length alone; energy enters as an additional factor.

BRIGHTNESS IN RELATION TO WAVE-LENGTH

Brightness, as we have seen, depends not only upon energy but also upon wave-length. Now, the very fact that radiant energy of wave-lengths above 800 or below 365 $m\mu$ is invisible to us implies such a dependence. No matter how great the energy of such waves, they produce no brightness at all. In other words, the *sensitivity* of the eye to radiant energy of these wave-lengths is zero.

Not only does the eye fail to respond to all wave-lengths, but also it is more sensitive to some of those waves lying within the limits than to others. In its unequal sensitivity for different wave-lengths the eye may be compared to a photographic plate, although the plate is especially sensitive to 'blue' and ultra-violet and only slightly affected by 'red' rays, whereas the retina is affected by red and is especially sensitive to yellow and green.

We may define the sensitivity of the eye to any given wave-length in terms of the quantity of radiant energy of that wave-length which is required in order to produce some standard degree of *brightness*. If waves of one length must have twice as much energy as waves of another length in order to produce the same degree of brightness, we say that the eye is only half as sensitive to the first kind of waves as to the second.

In order to determine the eye's relative sensitivity for different wave-lengths, we may have an observer compare each portion of the spectrum, in turn, with some fixed, standard stimulus which furnishes a constant brightness level. In each case the energy of the homogeneous stimulus is varied until the observer reports that the brightness produced by the homogeneous stimulus appears equal to the standard brightness. The energy, E, which the observer requires for each such brightness match is measured.[3] The results can then be represented in

[3] Radiant energy may be measured by means of an instrument known as the radiometer. In this instrument the energy is absorbed and converted into heat, and measured in terms of its heat equivalent.

the form of a *sensitivity curve*,[4] in which (since sensitivity is inversely proportional to E) we plot $1/E$ as a function of wave-length.

Now, as a matter of fact, it is found that the type of curve obtained depends on the *part of the retina* which receives the stimulus. Let us therefore consider first the simplest condition, in which the stimulus falls on the central part of the retina, in the region which contains cones and no rods. In this case the curve has the form shown by the full-line graph in Fig. 16.

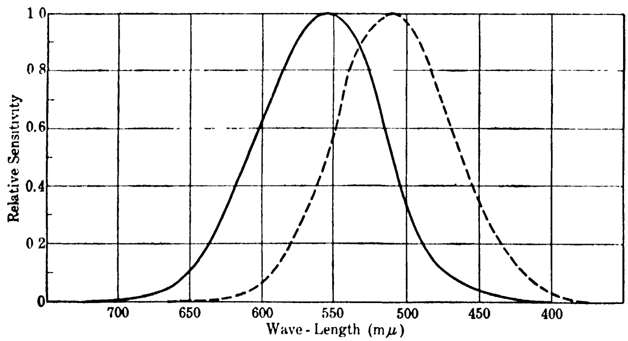


FIG. 16. SENSITIVITY CURVES.

Full-line graph = sensitivity curve of the cones (according to the standard data adopted by the International Commission on Illumination, 1931) Dashed-line graph = sensitivity curve of the rods. Data from Hecht and Williams.

The central retina is most sensitive at 554 mμ; this is a light towards the middle of the spectrum which is greenish yellow in hue. At wave-lengths which differ more and more from 554 mμ, the eye is less and less sensitive; and eventually the sensitivity falls to zero.

When the stimulus is confined to this rod-free area of the retina, the sensitivity curve has the same form no matter what the absolute brightness level may be. If two stimuli of different wave-lengths have the right energy ratio to produce equal brightnesses, they will continue to produce equal brightnesses, even though the energies of both are doubled, halved or changed in any proportion whatsoever.

[4] Often called the *visibility curve*.

When the part of the retina receiving the stimulus contains both *rods* and *cones*, we find a different state of affairs. The sensitivity curve does not have a constant form, but its shape depends on the strength or weakness of the stimuli which are used in determining it. If we make our brightness matches with stimuli of high energy, we obtain exactly the same curve as with the central rod-free area. If, however, we use very weak stimuli, we obtain a very different curve. This curve is represented by the dotted line graph in Fig. 16. (The data for this curve were obtained by measuring the energies required to match the various spectrum stimuli with the low brightness furnished by a surface of luminous paint.)

The curve as a whole has much the same form as before, but it is shifted in the direction of the shorter wave-lengths. The place of maximum sensitivity is no longer in the greenish yellow, but in the yellowish green part of the spectrum, at about 511 mμ. The eye has become relatively more sensitive to the short wave-lengths, in the blue and violet regions; at the same time it has become relatively less sensitive to the long wave-lengths in the red and orange part of the spectrum.

This change in the relative sensitivity of the eye for different wave-lengths which goes with a change in the general energy level is called the *Purkinje effect*, after the physiologist who discovered it in 1825. The Purkinje effect may be demonstrated by a very simple experiment. For this purpose one may use a piece of blue and a piece of red paper which, when viewed in normal daylight, are about equally bright (or the blue may be a little darker than the red). If now the illumination is reduced to a twilight level, the blue will become very decidedly brighter than the red. The difference in favor of the blue may continue to increase considerably as the eyes become 'adapted' (pp. 74-76) to the dim light. The fact is that, when the light-energy is weakened, the eye becomes more sensitive to the short waves which predominate in the light reflected by the blue paper, and less sensitive to the long waves which are predominantly reflected by the red paper.

In such an experiment, one notices another striking change in addition to that in the relative brightnesses. As the illumination is reduced, the colors become less and less *saturated*; and at the very lowest illuminations they lose their saturation completely. Instead of a blue and a red, one now sees a white or light gray, and a dark gray or black. At the lowest energies of light, all colors become *achromatic*—in other words, the eye becomes 'totally color-blind.'[5] As we shall learn later, these two parallel phenomena, the Purkinje effect and the loss of saturation of colors, are intimately connected.

The Purkinje effect does not occur in the central rod-free area of the retina. To obtain a rough verification of this fact, one may repeat the above experiment with blue and red objects of very small size, taking care to look directly at the objects, so that their retinal images stimulate the central region only. One then finds that the relative brightnesses of the blue and red are the same in weak as in strong light.

DARK-ADAPTATION AND THE SELF-LIGHT

When we pass suddenly from bright sunlight into a very dimly illuminated room, we can see nothing for a time until our eyes become *adapted* to the faint light. The retinas gradually become increasingly *sensitive*; they acquire the capacity to respond to quantities of energy which, at the first moment, were too small to stimulate them. This increase of sensitivity continues for half an hour or more—very rapidly at first, and then more slowly.

If, after having become thoroughly adapted to darkness or to dim light, we return to bright daylight, our eyes are now too sensitive for adequate vision under the new conditions, and we experience an annoying glare. Quickly, however, the sensitivity of the eyes diminishes, and the state of 'dark-adaptation' is replaced by one of 'light-adaptation.'

Thus the brightness resulting from a given stimulus is not

---

[5] 'Chromatically blind' would be a preferable term, for the eye still sees achromatic colors; but the other term is sanctioned by usage.

determined simply by the energy and wave-length of the light, but depends also upon the state of the retina. Brightness, like any other aspect of color, is a response to radiant energy rather than a mere reproduction of the properties of the stimulus itself; and there are many internal factors which can cause this response to vary even though the external stimulus remains the same.

As a matter of fact, we experience *brightness* even when there is no radiant energy stimulating the retina. When we go into a completely dark room, we do not get rid of visual experience. Instead, our whole visual field appears filled with an expanse of achromatic color—the so-called *self-light.*[6] The self-light represents an activity which is going on in the sensory system in the absence of any external stimuli. The faintest visible object furnishes a brightness which is just barely higher than that of the self-light.

The least strength of stimulus which will suffice to make a surface visible may be called the *light-threshold.* When we are light-adapted, this threshold is high; if we then become thoroughly dark-adapted, the threshold diminishes to 1/10,000 or even a smaller fraction of its original value. Conversely stated, sensitivity increases 10,000 times or more.

This increase in sensitivity, however, is not the same for all parts of the retina. The central *rod-free area* displays a relatively low capacity for dark-adaptation. If we test the eye with a very small stimulus which is carefully confined to the rod-free area, we find that the threshold decreases through adaptation only to about 1/100 of its initial value.

The periphery has a greater power of dark-adaptation than the center; this means that very faint objects which are invisible in central vision can be seen with the peripheral retina. When we try to observe a small, faint object, such as a dim star, we involuntarily tend to look a little to one side of it, instead of fixating it directly. In this situation the center of

[6] This phenomenon is sometimes called the 'self-light of the retina,' but there is evidence to show that it is due, at least in part, to spontaneous activity of visual processes in the brain.

the retina is no longer the region of best vision; and central fixation can be maintained only by a deliberate effort.

THE DUPLICITY THEORY

The peripheral retina and the central rod-free region, as we have seen, differ in two important respects. The Purkinje effect appears only in the periphery, not in the center. Furthermore, the periphery is superior to the center in its capacity for dark-adaptation. Let us also recall that in very dim light all vision is achromatic; this is another relevant fact, because extremely weak lights are seen only with the peripheral retina and not with the less sensitive central area.

These facts are all brought into relationship with one another by the *duplicity theory*, developed by the physiologist von Kries. According to this theory, the *rods* and *cones* of the retina differ in their functions. The rods have a much greater power of dark-adaptation than the cones, so that very weak lights can be seen only with the peripheral rods, not with the central cones. But these weak lights appear achromatic, because the rods are *totally color-blind*; color vision is a function of the cones alone.

The cones and rods differ in their distribution of sensitivity with respect to wave-length. The cones are maximally sensitive at 554 $m\mu$ (greenish yellow), and the rods at 511 $m\mu$ (yellowish green). The two curves of Fig. 16 are the sensitivity curves for cone-vision and rod-vision respectively. The Purkinje effect represents the transition from cone-vision to rod-vision—a transition which occurs only in the peripheral retina, where both rods and cones are present.

This theory is supported by the fact that nocturnal vertebrates have very few cones in the retina, whereas vertebrates which are active only in daylight possess very few rods. Moreover, there exist a few human beings who are totally color-blind even in bright daylight, and who are supposed to lack cones. This supposition is based upon the following facts: in the most frequent form of total color blindness the Purkinje

effect is absent; the sensitivity curve has its maximum at about 511 mμ at all levels of illumination; and in many such subjects the central retina is completely blind.

There are also certain persons who are congenitally 'night-blind,' or lacking in the normal power of adaptation to dim light. In these persons also the Purkinje effect is absent. The sensitivity curve has its maximum at about 554 mμ. This defect is attributed to an absence of rod-function.

The duplicity theory is so thoroughly supported by a wealth of different kinds of evidence that it may now well lay claim to the title of fact rather than theory.

THE PHOTOCHEMISTRY OF THE RETINA

Radiant energy acts on the photographic plate by causing chemical change in the materials with which it is coated—a so-called *photochemical reaction.* There are strong reasons for believing that the response of the rods and cones is a process of this same type.

The rods contain a bluish red pigment, the *visual purple,* which bleaches when it is exposed to light. If the bleached retina of an animal is kept in darkness for some time, its color is restored; an opposite chemical change evidently occurs which regenerates the original material. No photosensitive substance has yet been identified in the cones. It is probable, however, that the cones also contain such a substance, or a number of substances, but in very dilute form.

The first step in the visual process seems to consist, then, in the decomposition of photosensitive substance by the stimulus. The decomposition products, in turn, probably serve to instigate a second chemical reaction, involving other substances which are not directly sensitive to the light-stimulus; and the whole process which intervenes between the reception of radiant energy and the discharge of impulses in the optic nerve may well be very complicated. We can speak with more assurance about the first or photochemical stage than about these later stages.

If the visual response is due to chemical change in photosensitive substances, then it is logical to suppose that the reverse change, the regeneration of the original substances, furnishes the basis for dark-adaptation. The greater the amount of sensitive material which is present in the receptors at a given moment, the greater will be the amount of chemical change produced by a given quantity of energy—in other words, the greater will be the sensitivity of the retina. Exposure to a light-stimulus, on the other hand, will decrease the amount of available material for the photochemical reaction and thereby diminish the sensitivity. This chemical conception of the adaptation process has been recently worked out in mathematical detail by the physiologist Hecht (3).

The form of the *sensitivity curve* (Fig. 16) can also be simply explained in photochemical terms. The visual purple, like any other colored pigment, strongly absorbs radiant energy of certain particular wave-lengths. Now, the only radiant energy which can act on a photosensitive substance is that energy which is absorbed by the substance and converted into energy of chemical change. Energy which passes through the substance without being absorbed can produce no chemical effect.

It has been shown by Hecht that the sensitivity of the rods at any wave-length is proportional to the absorbing power of the visual purple at that particular wave-length. In other words, the sensitivity curve for rod-vision is simply an expression of the light-absorbing properties of the visual purple. It is entirely likely that the sensitivity curve of the cones, in an analogous fashion, represents the absorbing properties of a photosensitive substance (or group of substances) in these receptors.

STIMULUS MIXTURE

If we mix together two stimuli which produce different colors, what principles determine the color of the mixture?

Fig. 17 illustrates a scheme for the experimental study of

stimulus mixture. S_1 and S_2 are two differently colored areas. They may be, for instance, two areas of spectral light, each of a different wave-length, or they may be two areas of colored paper. A glass plate P is so placed that the light-rays from S_2 are reflected into the observer's eye, striking the same area of the retina as the direct rays from S_1.

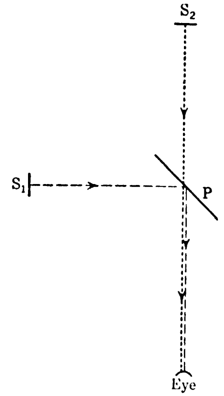


Fig. 17. Scheme for Experimentation on the Mixture of Stimuli.

The glass plate P reflects the light rays from S_1 into the observer's eye; the rays from S_2 pass directly through the glass and enter the eye in the same direction.

Another very convenient method for use with colored papers is the method of rotating disks. A disk made up of two differently colored sectors is rotated at such high speed that its surface looks perfectly uniform and shows no flicker. In this method the two kinds of reflected light do not actually strike the same area of the retina at the same time, but in very rapid alternation. However, the effect for perception is like that of a true light-mixture.

If we increase the size of one sector, and decrease that of the other, we are varying the relative amounts of *time* during which the two different stimuli act on any given retinal point. Under these special conditions of very brief exposure, our visual mechanism responds to a change in the *duration* of a stimulus in the same way that it would respond to a like change in the *energy* of the stimulus. Thus, if we enlarge one sector, at the same time reducing the other, this change is equivalent to increasing the energy of the first stimulus and diminishing that of the second. With these facts in mind, the results obtained with rotating disks can be validly compared with those obtained by true light-mixture.

The mixture of *paints* or pigments does not furnish a simple additive mixture of light-rays and is therefore unsuited for our experiments. Complicating physical factors enter into such a mixture.

Let us begin with the facts regarding *brightness*. The brightness of a color produced by mixture of stimuli is greater than the brightness which any of the individual stimuli could alone produce. This is a very familiar fact. Moreover, experimental evidence seems to show that the brightness of the mixture color depends only upon the brightnesses of the individual colors, and not upon their hues or saturations. For example, if we mix a yellow[7] with an equally bright blue, we obtain the same brightness that would be produced by mixing the yellow with an equally bright red, green or white.[8]

Now let us consider the principles governing the *hue* and *saturation* of a mixture color. Let us refer to the psychological 'hue circle' (Fig. 12). Consider any two stimuli which produce dissimilar hues. As a rule, the mixture of these stimuli yields a color which stands on the hue circle at a place *intermediate* between the positions of the component hues. Thus, for instance, the mixture of blue with green yields a bluish green or a greenish blue.

In general, moreover, the saturation of the mixture color is lower than that of either component; mixture involves a loss of saturation. This loss of saturation tends to be greater, the farther apart the two hues stand upon the hue circle.

The hue and saturation of the mixture color depend not only upon the hues and saturations of the components, but also upon their relative *brightnesses*. Consider, for example, the colors blue and green. Any increase in the brightness of the blue tends to make the mixture color more bluish; any increase in the brightness of the green shifts the mixture color towards green.

Suppose we begin by mixing two colors which stand rather near together upon the hue circle. The mixture hue will depend upon the brightness ratio between the components, but it will always be intermediate between these component hues. The saturation of the mixture will be low. Let us increase the

---

[7] It is often convenient to speak of 'mixing colors,' although, to speak precisely, we never mix colors; it is always *stimuli* that we mix.

[8] The accuracy of this law has been questioned (8, p. 550 f).

separation of the two hues and repeat the experiment. We shall obtain the same general sort of result as before, except that there will now be a greater loss of saturation.

If our two hues are sufficiently far apart upon the hue circle, this loss of saturation can be *complete.* In other words, the mixture no longer yields a chromatic color, but a gray or white. Two hues which produce such a result are said to be *complementary* to each other. Primary blue and primary yellow are one pair of complementary hues.

In the special, limiting case of complementary hues, mixture never results in an intermediate hue, but simply tends to bring about a loss of saturation. Thus, if a blue of high brightness is mixed with a yellow of low brightness, the mixture color is a blue of slightly diminished saturation. As more and more yellow is added, the saturation of the blue decreases further, and finally reaches zero (achromatic color). If still more yellow is added, the mixture color is a yellow, at first weakly saturated, and then more and more strongly saturated.

This example illustrates the typical results obtained when any two complementary hues are mixed. For every hue, there exists, at a rather remote point on the hue circle, a complementary hue. These two hues, when mixed in a certain brightness ratio, produce an achromatic color. When they are mixed in any other brightness ratio, the mixture color has the hue of the stronger component, in reduced saturation.

Although primary blue and primary yellow are complementary, primary red and green are not. The complementary to primary red is a bluish green, and the complementary to primary green is a purple. Thus hues which stand at diametrically opposite points of Fig. 12 are not, in general, complementary. We can, however, construct a new figure in which the positions of hues in Fig. 12 are so changed that complementary hues stand exactly opposite one another (Fig. 18).

Let us now agree to use Fig. 18 as a representation not only of the hues, but also of the saturation of colors. The center of the circle now stands for the zero point of saturation (an achromatic color). Distance from the center represents degree

of saturation, and the periphery of the circle stands for a series of maximally saturated hues, all of equal saturation (see Fig. 13).

With the aid of Fig. 18, we can summarize all the facts regarding hue and saturation of mixtures in terms of a very simple rule. This rule, though not rigorously exact, is sufficiently close to the truth for the purposes of an elementary discussion.

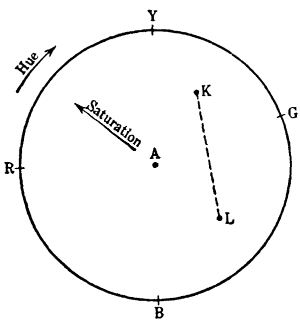


FIG. 18. FACTS OF STIMULUS MIXTURE.

R, Y, G and B stand for the four psychological primaries. *A* represents an achromatic color. *K* and *L* are two colors which are equal in saturation but have different hues.

Let the hue and saturation produced by any given stimulus be designated by an appropriate point in Fig. 18, say the point *K*. Let the hue and saturation of a second stimulus be represented by a second point *L*. Draw the straight line *KL*. Then *the hues and saturations of all the possible mixtures between the two stimuli are represented by all the various points in the line KL.* The exact location of any particular mixture color upon this line will depend on the brightness ratio between the individual colors. Increasing the relative brightness of *K* displaces the mixture color along this line in the direction of *K*; increasing the relative brightness of *L* displaces the mixture color towards *L*.

It is obvious that if we mix two hues which are equidistant from the center (or have equal saturations), the mixture color will stand nearer to the center than does either component (or will have reduced saturation). It is also easy to see that this reduction of saturation reaches its extreme in complementary hues. Our rule also implies that the mixture of noncomplementary hues furnishes an intermediate hue, whereas complementary hues tend only to reduce each other's saturation.

Let us consider the four primary hues, R, Y, G and B, and

note what happens when we mix two of these with each other. The mixture of R and Y yields a red-yellow (a yellowish red or reddish yellow). Similarly, Y and G mix to produce a yellow-green, G and B furnish a green-blue, B and R yield a blue-red. What happens with Y and B, or R and G? We cannot obtain a yellow-blue, or a red-green; no such hues exist.

Y and B, as we have already noted, are *complementary* to each other. If mixed in the appropriate brightness ratio, they yield a color which is quite different from either Y or B, namely a gray or white. R and G are not complementary. But (as Fig. 18 implies), if these hues are mixed in a certain definite brightness ratio, the result is an unsaturated *primary yellow*. In this case also, then, the mixture color is quite unlike either of the two primary hues which produce it. These facts are of much interest for the theory of color vision.

Another important fact can be readily deduced from Fig. 18. All the hues in the complete hue circle can be obtained by mixing only *three* properly chosen stimuli, in various proportions. Take, for example, three stimuli which individually furnish the primary hues R, G and B. By mixing the first two stimuli in different proportions of energy, we can produce all the hues lying between R and G in Fig. 18. Mixtures of the second two stimuli furnish all the hues lying between G and B. Finally, by mixing the first and third stimuli, we can obtain all the hues between B and R.

If we mix all three stimuli in a certain proportion, a gray or white will result. Mixtures of all three stimuli in other proportions will yield chromatic colors of relatively low saturation.

THE YOUNG-HELMHOLTZ THEORY OF COLOR VISION

The facts of stimulus mixture lead to a conclusion of prime importance for any theory of the physiological processes underlying color vision. On the one hand, the variety of physical stimuli which can act on the eye is almost infinite. It includes not only all the wave-lengths between 760 and 390 mμ, but

all possible combinations of any number of these wave-lengths in different energy proportions. On the other hand, if we consider the totality of possible *hues* which these stimuli can produce, we find no comparable variety. The sum total of hues which we experience, either as a result of homogeneous or of complex lights, is limited to four primary hues, red, yellow, green and blue, and their intermediates.

The physiological process upon which our experience of hue depends must, therefore, have fewer modes of variation than does the physical stimulus. Now, a very simple way of conceiving a process which has only a few modes of variation is to think of it as made up of a small number of individual part-processes or *elementary processes*, each of which can vary in degree. This idea has furnished the starting point for most theories of color vision.

The oldest and simplest theory of this type is one which was proposed by Helmholtz in 1852, following a basic idea outlined by Thomas Young in 1801. This theory still finds many adherents at the present time. We shall not attempt to follow its historical development, but shall proceed at once to state the theory in an up-to-date form.

·According to the Young-Helmholtz theory, there are three kinds of cones[9] in the retina. We shall speak of these as the R-cones, G-cones and B-cones. The R-cones, if they were stimulated in isolation, would give rise to the experience of red. Isolated stimulation of the G-cones would give rise to the experience of green, and isolated stimulation of the B-cones would give rise to the experience of blue. All colors, of whatever hue, saturation and brightness, result from the simultaneous stimulation of these three kinds of cones in different amounts and proportions.·

This theory provides a simple explanation for the facts of *stimulus mixture*. We have seen that by mixing three stimuli, which individually appear red, green and blue, one can produce a chromatic color of any desired hue, or an achromatic

[9] It is now generally agreed that the duplicity theory is correct in its contention that the rods play no part in color vision.

color. According to the Young-Helmholtz theory, the first stimulus acts predominantly on the R-cones, the second on the G-cones and the third on the B-cones. When one varies the proportions of red, green and blue in the mixture, one is varying the relative amounts of stimulation furnished to the three kinds of cones. The resulting hue depends upon the relationships between the three physiological processes.

If all three kinds of cones are responding at the same time, and in equal measure, we experience an *achromatic color.* In general, the more nearly equal the three responses, the more closely does the color resemble white—that is to say, the lower is its saturation.

Whenever a stimulus excites all three kinds of cones, but in unequal amounts, the resulting hue is determined by the two kinds of cones which are excited most strongly. For example, let us assume that a stimulus arouses responses R, G and B in the ratio 5: 7 : 11. The 5 units of R, together with 5 of the G-units and 5 of the B-units, go to the production of a certain amount of white. The only *chromatic* effect is that due to the excess 2 units of G and 6 units of B. The stimulus therefore appears greenish blue. The color is an *unsaturated* greenish blue, because of the admixture of 5 units each of R, G and B, which do not contribute to the hue but only tend to make the color look whitish. The *brightness* of the color depends on the total response of all the cones.

When the responses for R and G dominate over the B-response, one experiences yellowish hues. *Yellow* occupies a position in the theory which is analogous to that of white. It is not assigned a special receptor of its own, but it is assumed to correspond to *equality* of the R-response to the G-response.

The Young-Helmholtz theory finds favor with many contemporary physiologists, because it fits in very simply with prevailing views about the nature of the nerve impulse. It is widely believed at the present time that the impulses conducted by all nerve fibers are essentially alike. The impulses in the fibers of the optic nerve, for example, seem to be very

much like those in the sensory nerve fibers of the skin. The impulses in either kind of fiber vary in frequency with the strength of the stimulus. But no difference in the form of the impulses has been established which would account for the difference in sensory quality between visual and tactual experience. The physiologist infers, therefore, that this difference must be explained by reference to the brain—that it must depend in some way on the fact that the two kinds of fibers end at different places in the brain.

However, if the argument holds good as between vision and touch, why should it not also apply to the different qualities within a single sense, such as red and green in vision? If all impulses in the optic nerve are alike in kind, how can one explain hue discrimination except by assuming that the retina contains several different types of cones which have different connections with the brain? On such an hypothesis, a given stimulus excites the various types of cones in a characteristic proportion. Each type of cone sends impulses along its specific pathways to its specific place in the brain and presumably sets up a kind of brain process which is characteristic of that particular place. The hue which we experience is dependent on the ratio between the several kinds of brain process.

Of all theories of this type, that of Young and Helmholtz is the simplest possible, since it assumes only three types of cones. One may have a theory which postulates a larger number of types; the facts of stimulus mixture, from which the Young-Helmholtz theory takes its starting point, merely indicate that one must postulate at least three.

Let us next consider how, according to this theory, the responses of the three types of cones vary with the nature of the stimulus. Each kind of cone is assumed to be maximally sensitive to some particular wave-length, but to respond in less degree to other wave-lengths. Fig. 19 shows, for each species of cone, how the degree of response varies with the wave-length. According to these curves, most of the spectral stimuli excite all three groups of cones, but in different proportions.

It will readily be seen that the points *a, b* and *c* represent the three primary hues in the spectrum—yellow, green and blue. The hue and saturation of a *complex* stimulus can presumably be determined by summing up the effects of its individual wave-lengths on the R-cones, G-cones and B-cones, and then calculating the appropriate ratios.

These response curves are derived from the data of stimulus mixture, by a process which is too involved to be described here.

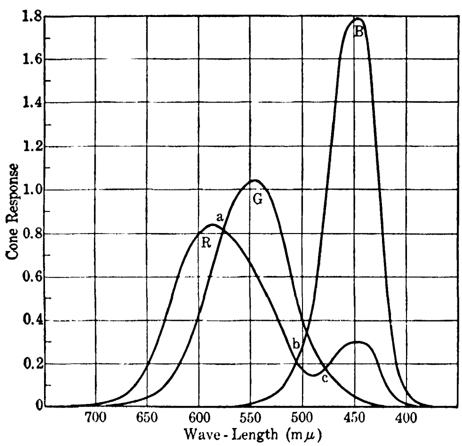


FIG. 19. R, G AND B CURVES OF THE YOUNG-HELMHOLTZ THEORY.

These curves represent the stimulus-mixture relationships for the normal observer (the 1931 standard observer of the International Commission on Illumination), and were constructed with the aid of Deane B. Judd The primary stimuli to which these curves refer have been chosen so that the following criteria are satisfied (1) the crossing points *a*, *b* and *c* correspond to the three psychological primaries of the spectrum—the primary yellow, green and blue—for the average observer (according to Westphal); (2) the G and B curves represent the stimulus-mixture relationships for one type of partially color-blind observer (the protanope); (3) the R and B curves represent the stimulus-mixture relationships for a second type of partially color-blind observer (the deuteranope).

Fig. 19 represents only one of a variety of forms in which the response curves have been plotted by different investigators. The data of stimulus mixture do not in themselves suffice to fix the precise form of the curves. Sometimes red, green and violet, rather than red, green and blue, have been assumed to be the colors corresponding to the three types of cones. How-

ever, the theory takes on a simpler and, to most critics, more plausible form when each variety of cone is associated with a primary color.

The Young-Helmholtz theory is nowadays seldom regarded by its most vigorous supporters as a basis for explaining all the phenomena of color vision. It is primarily an hypothesis about the function of the retinal cones and their connections with the brain. Such conceptions, if they are valid, would comprise only a relatively small part of a complete theory of color vision.

The Young-Helmholtz theory has been adversely criticized, especially by Hering and Ladd-Franklin, on the ground that it does not do justice to the properties of *white* and *yellow.* White does not appear to be a reddish greenish blue, and yellow does not seem like a reddish green. Since white and yellow are not intermediates between any other colors, but have an individuality of their own, it has been maintained that we have to assume some special kind of receptor (or some special kind of retinal process) for each of these colors.

In answer to this objection, it has been pointed out that our visual consciousness would seem to be determined directly by processes occurring in the brain cortex, and only indirectly by the response of the retina. The total color process in the brain is not necessarily a mere sum of three independent processes for R, G and B. When the retinal stimulus is such as to produce equal effects on these three brain processes, the total result may well be something unique, quite different from the processes for R, G or B occurring individually. Similarly, equal effects on the processes for R and G may give rise to another unique phenomenon, yellow.

Support of Helmholtz's theory is derived from the fact that the experience of white may be produced by simultaneously presenting a yellow stimulus to one eye and a blue stimulus to the other. Likewise, one can obtain yellow by the binocular mixture of red and green. In other words, white or yellow may be produced by a mixture which occurs in the brain—

for it is in the brain that the stimulations of the two eyes combine their effects. It would seem, therefore, that we do not need to assume special 'white' or 'yellow' receptors in the retina in order to account for the fact that white and yellow are unique colors.

INTERACTION OF VISUAL AREAS

The color which is seen at a given place in the visual field depends not only on the stimulation applied to the corresponding area of the retina, but also on the nature of the stimulation in *neighboring areas.*

One example of this principle is given by the phenomenon of *brightness contrast.* The brightness of any visual area is lowered by increasing the brightness of nearby areas, and conversely. Thus a small area of moderate light-intensity may appear white when it is presented on a dark background. If the surroundings are gradually made brighter, the brightness of the small area is continually decreased, and with sufficiently brilliant surroundings the original white may even be transformed into a black.

There is also a *chromatic contrast.* A small area of gray paper, when placed on a background of saturated blue paper, looks slightly yellowish. On a red background, the gray becomes bluish green. This effect is strongest if the gray and its background are about equally bright. It is also greatly enhanced by covering the colors with a sheet of tissue paper.

In general, any chromatic color tends to induce the *complementary* color in its neighborhood. Thus neighboring achromatic colors acquire a tinge of the complementary hue, and neighboring chromatic colors are changed in appearance. The more saturated a color, the greater is its chromatic effect on neighboring regions.

When two different colors are in proximity, each tends to change the other. For example, when two complementary colors adjoin each other, the saturation of each is increased by the contrast action of its neighbor. If red adjoins blue, the

red becomes yellowish, and the blue becomes greenish, by virtue of contrast.

The phenomenon of contrast implies that the physiological system underlying vision is not simply a group of independent local systems corresponding to the different parts of the visual field; it means that the different parts of the system can interact with one another. The physiological nature of this contrast interaction is unknown, but there is evidence to indicate that the interaction occurs not in the retina, but in the brain.

One line of evidence is furnished by experiments on the interaction of two colors which are separately presented to opposite eyes (binocular contrast). In this case no interaction can take place except in the brain. According to the experiments of Köllner (6), the contrast-effects obtained under these conditions are fully as strong as when both stimuli are seen by the same eye. This result implies that the interaction underlying contrast is wholly confined to the brain, and that the retina plays no part in it. Further data point to the cerebral cortex as the locus of the interaction. For instance, Müller (9) has reported a patient having a cortical injury who displayed almost a complete loss of contrast phenomena.

In the so-called phenomena of *summation* we have to do with a sort of interaction which is quite different from contrast. It has been found, in experiments on the light-threshold, that the visibility of a weakly illuminated area depends on the *size* of the area, and not merely on the strength of the illumination. Thus, a very minute area may be invisible, whereas a larger area which receives the same illumination may be easily perceptible. In other words, the visibility of the stimulus is not determined simply by the amount of energy delivered to each receptor in unit time, but varies with the *number of receptors* which are responding. With very small stimuli, any increase in area entails a decrease in the strength of light required to make the area barely perceptible; but eventually a certain size is reached beyond which there is no further improvement of visibility. This dependence of visibility on the area of the retinal

image is most pronounced in the periphery of the visual field, and especially in the extreme periphery; but it is not absent in the center.

Under the conditions of this experiment there appears to be a mutual *reinforcement* of the responses in neighboring regions —a so-called *brightness summation.* Such a reinforcement occurs not only between the individual parts of a single stimulated area, but also between small separated areas which are not too far apart. Thus, two small areas, one near the other, may be visible, although either area alone would be invisible.

The retina, it will be recalled, not only contains neurons for the conduction of impulses away from the rods and cones, but it also contains laterally placed neurons which connect one part of the retina with another. Such interconnecting pathways are typical of nervous centers rather than of sense organs, and they are found in no sense organ except the eye. Recent experiments by Adrian and Matthews (1) have shown that these retinal interconnections are concerned in the process of brightness summation. By means of the lateral neurons a response set up at any point in the retina can reinforce the responses at neighboring points. The 'summation' of spatially separated stimuli is a characteristic property of nervous centers; and the retina, in function as well as in structure, is a true nervous center and not merely an organ for the reception of stimuli.

Besides brightness summation, there is also a *chromatic summation.* A chromatic color tends toward gray when its area is made very small. In other words, color saturation—within certain limits—increases with area. Neighboring areas which receive the same kind of stimulation evidently reinforce each other in chromatic effect. Chromatic summation, like brightness summation, is more pronounced in the periphery of the visual field than in the center.

In contrast and summation we have two different types of interaction, each of which has definite biological advantages. Contrast tends to sharpen the contours of objects and to accentuate their differences of brightness and color. Summation,

on the other hand, facilitates the perception of small and weak stimuli.

BLACK

Black occupies a very peculiar position among the colors. Popularly, black is regarded as corresponding to the absence of stimulation. But the 'self-light' which we experience in the absence of any external stimulation is grayish and not black. However, an unlighted area which is *surrounded* by an area of strong light may appear, through the action of contrast, in a brightness which is much lower than that of self-light. Even a lighted area may look black, if its surroundings are sufficiently intense.

Black may also be produced as an *after-effect* of strong stimulation (an after-image, pp. 95f.). Thus, when we pass from strong light into a dark room, our visual field appears black rather than gray during the first few moments.

In both these cases, black results from *indirect action of intense stimulation*, rather than from the simple absence of stimulation.

In our everyday perceptions of black *objects* (surface colors), another and a more complicated set of factors comes into play. Here we must take into account the distinction which perception makes between brightness of *surface-color* and brightness of *illumination.* A table, or a piece of cloth, which looks black in noon sunlight also looks black in twilight. In the one case, however, it appears as a black surface which is strongly illuminated, and in the other as the same black surface weakly illuminated.

The fact that an object displays a black surface-color does not necessarily mean that the object reflects only a small quantity of light into our eyes. If a black object is strongly illuminated, it may reflect a large quantity of light. But the light reflected by a black-appearing object is always *relatively* weak —weak, that is, *in relation to the illumination.* Similarly, a white object may reflect only a small amount of light, if the

illumination on it is weak. But the light it reflects is always *relatively* strong, and stronger than that reflected by any gray or black object which lies *in the same illumination.*

In general, then, our perception of a surface-color as black, white or gray is not determined simply by the quantity of light which the surface reflects into the eye, but involves the total situation in which the color is seen. From this total situation we receive not merely the impression of so many particular objects, but also the impression that these various objects lie in a certain kind of general illumination. Now, it looks very much as if a reasoning process might be involved in our perception of surface-colors. It is as if, in the case of each object, we took into account two factors: (1) the amount of light reflected by the object (the retinal stimulation it furnishes) and (2) the apparent illumination, as if we then inferred from the relationship between these two factors, that the given object ought to appear white, black or gray as the case may be. However, there can be no doubt that the process, the 'taking account of' the total situation, is a matter of immediate perception and not of reasoning.

The problem of the nature of this process is an interesting and important one for the theory of perception and will be discussed later (pp. 276f.). For our present purpose, it is sufficient to note that the perception of surface-colors is a complex affair, in which impressions of illumination seem to play a fundamental part. In particular, a surface-color tends to appear *black* when the light its object reflects is weak in relation to the illumination falling upon the object.

LIGHT-ADAPTATION; CHROMATIC ADAPTATION

Light-adaptation is the converse of dark-adaptation (pp. 74f.). Light, acting upon the eye, reduces its sensitivity; and the stronger the light, the greater the loss of sensitivity.

Suppose that one looks steadily, for a period of a few minutes, at a small bright area on a dark ground. As time goes on, the retina responds less and less vigorously to the stimulating

light, and the area appears less and less bright. The sensitivity continues to decrease for several minutes after the first moment of fixation, and thereafter remains at a constant, low level.

This loss of sensitivity becomes very obvious if one suddenly introduces a comparison stimulus, exactly like the one which is being fixated, at some nearby place in the dark ground. The comparison stimulus, acting on a retinal region which has not been desensitized by strong light, looks brighter than the fixated stimulus.

Under the ordinary conditions of vision, the eyes are constantly moving, and thus no part of the retina has time to become completely adapted to any stimulus. Complete adaptation is seldom reached except under artificial conditions of steady fixation.

If a *chromatic* stimulus acts continuously on the eye, its color decreases not only in brightness but also in saturation. A color which is not, in the first place, too highly saturated, or too bright, loses its saturation completely after a few minutes of adaptation and appears gray. With very saturated and very bright stimuli, chromatic adaptation does not reach this extreme; and, when the eyes are moving, the loss of saturation is relatively slight.

Let us see how chromatic adaptation may be interpreted by the Young-Helmholtz theory. Consider, for example, a primary green stimulus, which evokes, we shall say, a response of magnitude 9 on the part of the G-cones, and equal responses of magnitude 4 on the part of the R-cones and B-cones. All three kinds of cones will become steadily less sensitive, and give weaker responses to the stimulus as it continues to act. The decrease in sensitivity will be greater for the more strongly stimulated G-cones than for the R-cones and B-cones. The G-reponse is lowered, let us say, by 2/3, while the other two responses are lowered by only ½. The final result, then, is a G-response of magnitude 3, and an R-response and B-response each of magnitude 2. The ratio of responses, originally 9 : 4 : 4, is now 3 : 2 : 2. Adaptation has brought the ratio nearer to the

1 : 1 : 1 value which represents achromatic color. In other words, adaptation has made the color less saturated.

AFTER-IMAGES

Let the reader look steadily at Fig. 20 for about 40 sec., with unwavering fixation on some particular point in the figure, and then suddenly transfer his gaze to a blank sheet of white paper. On the paper will appear a *negative after-image*, an illusory figure which is like the original stimulus in form, but with its brightness relations reversed. The original blacks are replaced by whites, and the whites by blacks, as in a photographic negative.

FIG. 20. FIGURE FOR THE DEMONSTRATION OF THE NEGATIVE AFTER-IMAGE.
After A. Noll. Psychol. Forsch., 1926, 8, 7.

If one considers the facts of light-adaptation which we have discussed, it will become clear that just such an after-effect might be expected to appear under the conditions of the experiment. While the figure was being fixated, certain parts of the retina were exposed to the strong stimulus provided by the white patches in the figure. We should expect these retinal areas to become less sensitive to light than the neighboring areas on which the black patches formed their retinal images. This difference in sensitivity would become more and more pronounced with continuing fixation. What ought to happen when a homogeneous area of white paper is substituted for the original stimulus? The light reflected by the paper should evoke a strong response from the more sensitive, a weak response from the less sensitive, retinal areas. In other words, we should expect a pattern like the figure, but with opposite brightness relations, to appear on the paper.

If the eye is adapted to a *chromatic* stimulus by continuous fixation, and is then exposed to a uniform white surface, the after-image tends to be complementary in hue to the original stimulus. By Helmholtz's theory, the *complementary after-image* is due to the fact that the cones which are most strongly excited by the stimulus suffer a greater reduction in sensitivity than the less strongly excited cones. Thus, if the eye is adapted to green light, the G-cones become less sensitive than the R-cones and the B-cones. If, now, the eye is exposed to a white surface, the three kinds of cones do not, in their normal fashion, respond in equal measure, but the R-response and the B-response predominate over the G-response. Thus the after-image is complementary in hue to the adapting color.

Besides these after-images of 'negative' brightness and complementary hue, there are other after-images which resemble the original stimulus in brightness (*positive after-images*) and in hue (*homochromatic after-images*). For example, a glowing match, whirled about in the dark, is followed by a positive after-image in the shape of a long bright streak. Positive after-images are most easily obtained from strong stimuli which act only momentarily. Positive and homochromatic after-images are probably due to the persistence of a state of excitation in the retina, so that impulses continue to be discharged in the optic nerve after the removal of the stimulus.

PERIPHERAL VISION

Peripheral vision is less distinct than vision with the central retina (p. 215). The periphery also differs from the center in that it has a more primitive type of *color vision.*

Under ordinary conditions, these differences between direct and indirect vision do not impress us. The eye is in constant motion. Different parts of the visual field are viewed in quick succession with the central retina, and the total effect for perception is that of a field which is fairly distinct throughout its whole extent, each part possessing the color in which it appeared to central vision. However, if one keeps the eye steadily

directed at some given point, and meanwhile carefully observes the color of an object lying above, below or to one side of the fixation point, the peculiar color vision of the periphery becomes manifest.

When a stimulus is moved from the center towards the extreme periphery, its color decreases continually in *saturation.* Stimuli of moderate intensity, such as are furnished by colored papers, lose their saturation completely in the extreme periphery and appear *achromatic.*

For most colors, this loss of saturation is accompanied by a change of *hue.* As a stimulus passes away from the center, its color tends very quickly to lose any *reddishness* or *greenishness* which it originally possessed, while any *yellowishness* or *bluishness* which originally belonged to it is retained. Thus, an orange changes towards yellow before it finally becomes gray, and a purple changes towards blue.

Hering suggested that the retina may be regarded as containing three zones, each displaying a different type of color vision. In the central zone our vision includes reds, greens, yellows and blues, in addition to achromatic colors; whereas in the middle zone, according to Hering, experience is confined to yellows, blues and achromatic colors, and in the outer zone all stimuli appear achromatic.

However, experiment shows that these zones have no sharply defined limits. The limits which one experimentally finds for a given zone are always relative to the particular conditions under which the test was made. For example, a stimulus which has a weakly saturated color (in central vision) becomes achromatic at a small distance from the center. A stimulus of high saturation, on the other hand, may not become achromatic until the extreme periphery is reached.

The experimental limits of the zones also depend on the brightness of the test colors. If very strong stimuli are used, chromatic colors can be seen even in the extreme periphery. The color response of a given peripheral region also depends on the size of the stimuli and the brightness of the background on which they appear. With colors of large area, presented on

a background of equal brightness, the dissimilarity between peripheral and central vision is least marked.

Under any stimulating conditions, however, we find characteristic differences between color vision in the center and the periphery of the retina, which become more pronounced as the extreme periphery is approached. There is no satisfactory explanation of them.

It might be thought that the loss of saturation in the periphery could be attributed to the presence of a certain proportion of rods in that part of the retina—a proportion which becomes greater with increasing distance from the center. However, the rods play a significant role only when the retinal stimulus is weak,[10] while the loss of saturation occurs even with the very intense lights. Evidently, then, color vision for the peripheral cones is different from that for the central cones. This difference reaches its extreme in the cones at the edge of the retina.

COLOR BLINDNESS

Color blindness may be either congenital or acquired, and it may also be either total or partial. A totally color-blind person is one who lacks hue discrimination completely. A partially color-blind person possesses hue discrimination but frequently fails to distinguish between stimuli which appear dissimilar to the normal person. Acquired color blindness is caused by injury or disease of the nervous structures concerned with vision; it may be either total or partial.

Partial color blindness is a very frequent congenital abnormality. It is displayed by about 3 per cent of the male and 0.1 per cent of the female population. The commonest variety of partial color blindness is often called 'red-green blindness,' because persons with this defect characteristically fail to distin-

---

[10] When we pass from cone-vision to rod-vision, there is a change in the relative *brightnesses* of different colored lights—the Purkinje effect (pp. 73f.). When we pass from central to peripheral vision, there is no Purkinje effect except with relatively weak lights. Peripheral vision at the higher intensities, therefore, is cone-vision and not rod-vision.

guish between the colors of objects which, for the normal person, are colored red and green.

There are, however, two distinct types of red-green blindness, called *protanopia* and *deuteranopia.* The protanopes[11] display a striking abnormality of brightness perception as well as one of hue discrimination. Their sensitivity curve (p. 72) is quite different from the normal; they are extremely insensitive to red light, and the spectrum, to them, appears shorter than to the normal person. The brightness sensitivity of the deuteranopes, on the contrary, does not differ greatly from the normal.

The laws of stimulus mixture for protanopes and deuteranopes are different from those applying to normal persons. For a color-blind individual of either type, any light in the spectrum can be exactly matched by a suitable mixture of *two* lights, one from the long-wave end and the other from the short-wave end of the spectrum.[12] The protanopes and deuteranopes differ from each other in the relative amounts of long-wave and short-wave light which they use in making any such match.

If we compare the color discrimination of the red-green blind with that of normal persons, we find both a certain dissimilarity and a certain resemblance. On the one hand, the color-blind person is often unable to distinguish between stimuli which have different colors for the normal person. On the other hand, any stimuli which look exactly *alike* to the normal person also look exactly alike to both protanopes and deuteranopes.

Thus, although the protanopes and deuteranopes often reject each other's color matches, they both accept any color match made by a normal observer. These facts suggest that red-green blindness represents, in some sense, a simplification of normal color vision, rather than a totally different type of perception.

---

[11] The partially color-blind persons who have the abnormality called protanopia or deuteranopia are called respectively *protanopes* or *deuteranopes.*

[12] For this reason partially color-blind persons, both protanopes and deuteranopes, are sometimes called *dichromats*; whereas normal persons, who require three lights to match the whole spectrum, are called *trichromats.*

However, in protanopia and deuteranopia we evidently have two different kinds of simplified color systems.

Helmholtz attributed partial color blindness to a lack of function of one of the three types of cones in the retina. Protanopia, according to him, is due to a lack of R-response, and deuteranopia to a lack of G-response. The responses of the protanopic retina would then be represented by the G-curve and the B-curve of Fig. 19 (p. 87), and those of the deuteranopic retina by the R-curve and the B-curve. This assumption provides a very simple explanation for all the various facts we have mentioned so far.

According to Helmholtz's hypothesis, any stimulus acting on the retina of the protanope should give him the experience of green, blue or of some intermediate between these two hues. Lights which look white to the normal person should appear blue-green to the protanope. Similarly, the color world of the deuteranope should be confined to red, blue and their intermediates. The normal person's white light should appear purple to the deuteranope.

How do colors appear to the partially color blind? A study of their color matches does not answer this question. We learn, for example, that certain stimuli which look red and green to the normal person have no color difference for the red-green blind, but we do not thereby learn what the abnormal person's color experience is like. There are, however, certain individuals who are color-blind in one eye and possess normal color vision in the other. By comparing the colors seen with two eyes, such persons can relate the visual world of the color-blind to the world of normal experience.

On the basis of our rather scanty information about such cases, the following tentative conclusions can be drawn. The protanopes and deuteranopes do not lack the experience of white, as Helmholtz supposed. Stimuli which appear white in normal vision also appear white in abnormal vision. In addition to white, the color world of both protanopes and deuteranopes includes the two hues *yellow* and *blue*, in different gradations of saturation. The long-wave parts of the spectrum appear yellow; the short-wave parts appear blue.

Thus, although Helmholtz's hypothesis accounts for the color matches of the protanopes and deuteranopes, it fails to account for the nature of their color experiences. Various modifications of the hypothesis have been proposed in order to remedy this deficiency. Thus, it has been suggested that red-green blindness may involve abnormalities in the nervous structures of vision as well as in the retina.

REFERENCES

1. ADRIAN, E. D., and MATTHEWS, R. The action of light on the eye. III. The interaction of retinal neurones. *J. Physiol.*, 1928, 65, 273-298.
2. GRAHAM, C. H. Vision: III. Some neural correlations. In *Handbook of general experimental psychology*. Worcester, Mass.: Clark University Press, 1934. Pp. 829-879.
3. HECHT, S. Vision. II. The nature of the photoreceptor process. In *Handbook of general experimental psychology*. Pp. 704-828.
4. HECHT, S. The retinal processes concerned with visual acuity and color vision. Howe Lab. Ophth., *Harvard Med. Sch. Bull.*, no. 4, 1931.
5. HELMHOLTZ, H. L. F. VON. *Treatise on physiological optics* (translated from the third German edition). Rochester, N. Y.: The Optical Society of America, 1924-25. 3 vols.
6. KOLLNER, H. Ueber die Lokalisation des Simultankontrastes (und der gleichsinnigen Induktion) innerhalb der Sehbahn. *Arch. f. Augenheilk.*, 1916, 80, 63-103.
7. LADD-FRANKLIN, C. *Colour and colour theories*. New York: Harcourt, Brace, 1929.
8. McDOUGALL, W. Some new observations in support of Thomas Young's theory of light- and colour-vision. *Mind, N. S.*, 1901, 10, 52-97, 210-245, 347-382.
9. MULLER, G. E. Ueber die Farbenempfindungen. *Zsch. f. Psychol.*, 1930, *Ergbde.* 17 and 18.
10. PARSONS, J. H. *An introduction to the study of colour vision*. Second edition. Cambridge: Cambridge University Press, 1924.
11. RIVERS, W. H. R. Vision. In E. A. Schafer's *Text-book of physiology*. London: Pentland, 1900. II, pp. 1026-1135.
12. TROLAND, L. T. Vision: I. Visual phenomena and their stimulus correlations. In *Handbook of general experimental psychology*. Worcester, Mass.: Clark University Press, 1934. Pp. 653-703.
13. TROLAND, L. T. *The principles of psychophysiology*. New York: D. Van Nostrand Co., 1930. II, 51-205.
14. [Various.] *Report of a joint discussion on vision held on June 3, 1932, at the Imperial College of Science by the Physical and Optical Societies.* London: Physical Society, 1932.

Chapter 5

AUDITION

An understanding of the sense of hearing, as is true of vision, involves a knowledge of the physics of the stimulus, the structure and operation of the receptor and the nature of the phenomena which arise when the receptor is stimulated. In this chapter, therefore, the first three divisions will be (1) the physical nature of sound, (2) the anatomy and physiology of the ear and (3) the phenomena of hearing. A final division, (4) auditory theory, will include an attempt to coordinate our knowledge of the above three phases of the process of hearing.

THE PHYSICAL NATURE OF SOUND

Many objects, when energy is applied to them in an appropriate way, are set into a state of vibratory movement. A bell when struck, or a violin string when bowed, will present a long series of back-and-forth movements, continuing until the imparted energy is exhausted.

If the body is vibrating in air, as is usual, a part of the energy is transmitted from the moving surfaces to the air particles in the vicinity. And further, through a series of collisions between the air particles themselves, the vibration is conducted outward as waves of sound to all parts of the surrounding air.

One type of vibratory movement is of fundamental importance. This is *simple harmonic motion*, the simplest kind of vibration known. Not many bodies give this kind of vibration exactly, but many approximate it. The most familiar instrument that gives simple harmonic motion in nearly pure form is the tuning-fork.

The simple pendulum also executes simple harmonic motion,

This chapter was written by E. G. Wever of Princeton University.

though it is too slow in its movements to produce sound. The slowness of the movements is an advantage to their study, for it permits direct visual examination of them. If a pendulum bob, as in Fig. 21, is displaced from its position of rest and then released, it will swing regularly over a definite path. Each to-and-fro movement, from *L* to *R* and back again to *L*, constitutes one complete vibration, or one cycle, for which we use the sign ~. The time required for one cycle is the period of the vibration; the reciprocal of this, the number of cycles in one second, is the frequency. The distance *LM*, the maximum displacement from the position of rest, is the amplitude of the vibration. It will be noted that, as the pendulum uses up the energy originally imparted to it, the amplitude gradually diminishes, but the period remains constant. The same is true of all such vibrating bodies, and can easily be observed in the

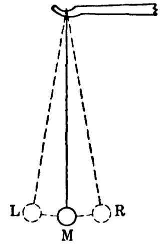


FIG. 21. THE SIMPLE PENDULUM.

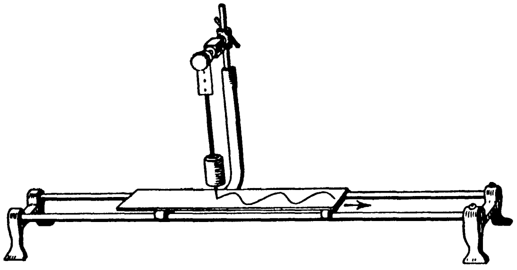


FIG. 22. METHOD OF RECORDING THE VIBRATIONS OF A PENDULUM.
The smoked glass plate is drawn from left to right while the pendulum is swinging to and fro.

tuning-fork, whose tone retains its particular quality while its vibrations die away in strength.

The details of vibratory motion are conveniently studied in a graphic record. Such a record for the movements of a pendulum may be obtained by the method illustrated in Fig. 22. A smoked-glass plate is drawn uniformly to the right, while a

pendulum fitted with a marking point is moving back and forth. The result is that the back-and-forth motions are drawn out in an undulating curve, which is known mathematically as a sine curve. Such a curve is shown in enlarged form in Fig. 23.

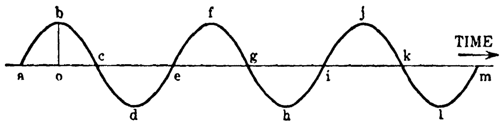


FIG. 23. A SINE CURVE.

The curve is drawn as vertical displacements from the time line *am*. One cycle is the portion *abcde*, or *bcdef*, or any similarly recurrent portion regardless of the starting point. The distance *ob* is the amplitude. The time elapsing from *a* to *c* is one period.

Simple harmonic motion has a very elementary relation to uniform circular motion. If we observe the point *P* in Fig. 24, which is considered as moving uniformly in a circular path as indicated by the arrow, and measure from time to time the vertical distances *PO* from the point to a fixed horizontal line *XX′*, these distances will vary in the form of a sine curve. This relation is easily observed in the driving mechanism of a locomotive. The driving wheel executes regular circular motion, while the driving rod moves back and forth in simple harmonic motion.

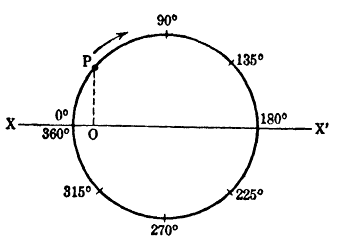


FIG. 24. THE RELATION BETWEEN CIRCULAR AND HARMONIC MOTION.

As *P* moves uniformly around the circle, the displacement *PO* changes harmonically.

It is sometimes necessary to state the position of a vibrating body exactly, as when for example we wish to indicate the point on its swing that a pendulum bob occupies at a particular instant. We require a system of measurement by which positions within any period of a vibration can be designated. The direct relation between harmonic and circular motion makes it convenient to transfer to harmonic motion the familiar system of circular measurement. We divide the circle into 360 degrees, by which we can indicate

any position. Likewise, we may divide the period of vibratory motion into 360 equal parts or degrees, and thus designate any position. This positional feature of vibration is called the *phase*, and the relation between two different vibrations when position in the cycle is considered is their phase relation. In Fig. 24 the circle is marked off in degrees, and it may be seen that, when point *P* is in the 0°, 180° or 360° positions, the displacement *PO* is zero, when it is at 90° or 270° the displacement is a maximum, and at other points the displacement is of intermediate value. If we refer to Fig. 23, it will be seen that point *a* corresponds to the zero phase position, point *b* to the 90° phase position, and so on. The point *e* has the same phase as *a*, the

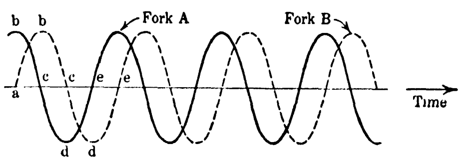


FIG. 25. THE 90° PHASE RELATION.
Fork *A* leads fork *B* in phase by 90°.

point *f* the same phase as *b* and, in general, corresponding points of different cycles have identical phase positions.

Two vibrations otherwise similar may differ in phase; thus if we strike two identical tuning-forks, one slightly before the other, we obtain movements as in Fig. 25, where corresponding points in the waves are lettered similarly. It will be observed that fork *B* is in the *a* or zero phase position when fork *A* has advanced to the *b* or 90° position, and when an instant later *A* is at *c* (180°), *B* is at *b* (90°), and so on. At every moment during the progression of the waves, the movements of fork *A* are ahead of those of *B* by 90°, and hence we say that fork *A* leads in phase by 90°. It is the same thing to say that fork *B* lags in phase by 90°.

If the two forks are struck exactly simultaneously, their movements will be in phase, that is to say, there will be no

phase difference. If fork *B* is struck just half a period later than *A* the phase difference will be 180°, which is called contrary

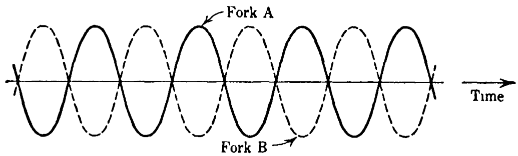


FIG. 26. THE OPPOSITE OR 180° PHASE RELATION.

or opposite phase. This last condition of contrary phase, in which the two movements are precisely opposite at all times, is shown in Fig. 26. Two vibrations may have any phase relation between exact agreement and complete opposition.

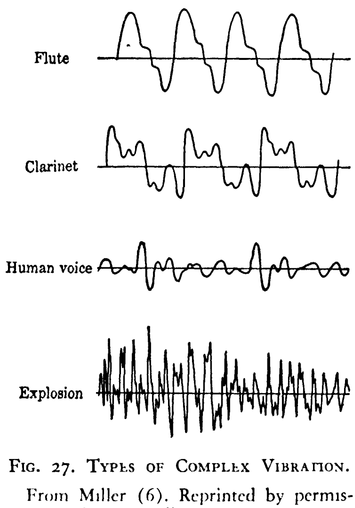


FIG. 27. TYPES OF COMPLEX VIBRATION.
From Miller (6). Reprinted by permission of the Macmillan Co.

Most vibratory movements are not so simple as those of the tuning-fork and pendulum, but have forms of various complexity. Several types of complex vibration are shown in Fig. 27. The first three curves represent the waves of a flute, a clarinet and the singing voice; the fourth represents the explosion of a sky-rocket (6). Of these the first three are periodic, since the wave form is repeated at least several times; the last is aperiodic, with no exact repetitions. We shall study later the relation between wave form and auditory experience.

THE ANATOMY AND PHYSIOLOGY OF THE EAR

The ear may be considered anatomically as divided into outer ear, middle ear and inner ear. The outer ear includes the

visible portions, an expanded flap called the *auricle* or *pinna*, and a short tube, the external auditory canal. The auricle in man has no important function and may be lost without appreciable impairment of hearing. The auditory canal is the path of entrance of aerial waves and leads internally to the eardrum, which forms the boundary of the middle ear.

The middle ear is contained within an irregularly shaped cavity in the temporal bone. The cavity is filled with air, conveyed by the *Eustachian tube* which enters from the throat.

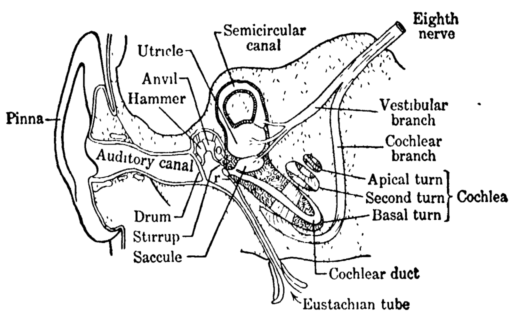


FIG. 28. DIAGRAM OF THE EAR.
The stippled regions represent bone, and the shaded portions indicate cavities filled with perilymph. *r*, round window. *o*, oval window. After Czermak.

The Eustachian tube is normally closed, but opens during the act of swallowing and permits an adjustment of the pressure of the air within the middle ear and that of the atmosphere.

A chain of three small bones, the *auditory ossicles*, leads across the cavity of the middle ear from the drum to a small oval-shaped membrane, the *oval window*, which closes an opening in the inner wall of the cavity. The first ossicle, the *hammer*, is attached to the inside surface of the drum; the third, the *stirrup*, nearly covers the oval window to which it is attached by a ligament; the middle ossicle, the *anvil*, connects these two ossicles. A number of ligaments hold the ossicular chain together and keep it suspended in the cavity of the middle ear; two of these ligaments are attached to muscles whose contractions control the tensions of the drum and ossicles.

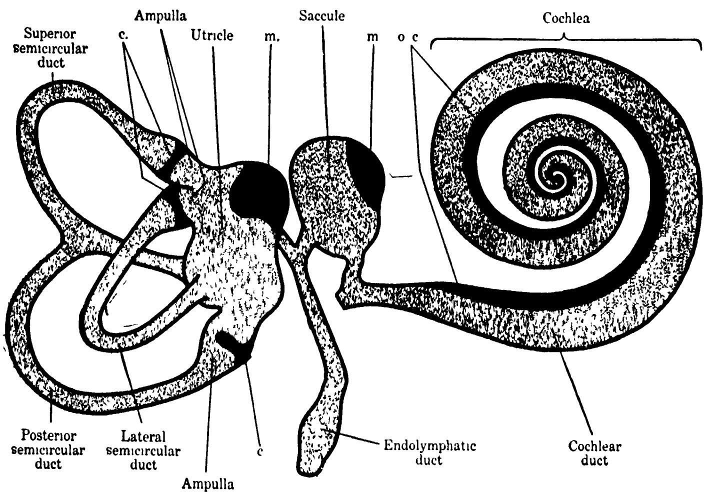
When sound waves enter the auditory canal and strike upon the eardrum they set it in motion, and thereby move the arm of the hammer which is attached to the drum. The movements are transmitted through the anvil to the stirrup and finally to the fluid content of the inner ear by way of the oval window against which the stirrup rests.

The middle-ear system is of considerable importance in determining the sensitivity of the ear as a receiver of air vibrations, because it acts as a kind of mechanical transformer. Vibratory energy is not readily transferred directly from a light medium like the air to a denser medium like the fluid of the inner ear, but through the middle-ear system the transfer is greatly facilitated. The transformer action occurs in two ways. Firstly, the eardrum is about twenty times larger than the oval window, which gives a corresponding increase in pressure. Secondly, the ossicles are arranged in the form of a lever system with a reduction ratio of 1.5 to 1, and therefore give an increase of pressure of 1.5 times. These two effects together produce an increase of pressure of 20 x 1.5 or 30 times, and hence a very efficient transmission of the sound energy.

The apparatus of the middle ear does not act as a perfect mechanical system, especially when the sounds are loud. The reason lies partly in the peculiar properties of the eardrum, which does not respond equally well to all sounds; and partly in the nature of the joints of the ossicles. These joints have a small amount of looseness or 'play' in them, and therefore allow a degree of discrepancy between the drum movements and the movements impressed upon the stirrup. The situation may be compared to that in a simple lever system where the arms of the lever are subject to a small amount of bending. These effects, which cause the movements of the fluid of the inner ear to depart from a strict proportionality to the air waves, constitute middle-ear distortion and give rise to a number of peculiar auditory phenomena, which will be described later.

The inner ear is contained in the bony labyrinth, a space of complex form hollowed out in the deeper portion of the tem-

poral bone. Three divisions of the labyrinth are recognized: the *vestibule*, *cochlea* and *semicircular canals*. The cavities of these parts contain a fluid, the *perilymph*, in which lies a complex system of membranous sacs, the membranous labyrinth. Of the several parts of the membranous labyrinth, only the cochlea has to do with hearing. The semicircular ducts, utricle and saccule are concerned with bodily equilibrium.



Black areas are those with nerve structures
m. = macula c = crista
o c. = organs of Corti

FIG. 29 DIAGRAM OF THE MEMBRANOUS LABYRINTH.

The cochlear duct is a spiral-shaped tube lying to one side of the spiral bony canal of the cochlea. Like other parts of the membranous labyrinth, it contains a fluid, the *endolymph*, which is distinct from the perilymph outside the duct. As seen in cross-section in Fig. 30, the cochlear duct is of triangular form, and lies between two larger tubes, the vestibular canal and the tympanic canal, from which it is separated by membranes. The *vestibular* and *tympanic* canals extend to the basal end of the cochlea at the vestibule, where each communicates with the cavity of the middle ear through a perforation in the

lateral wall of the vestibule; these perforations are the *oval* and *round windows.* The first, the oval window, which opens from the end of the vestibular canal, bears a membrane to which the

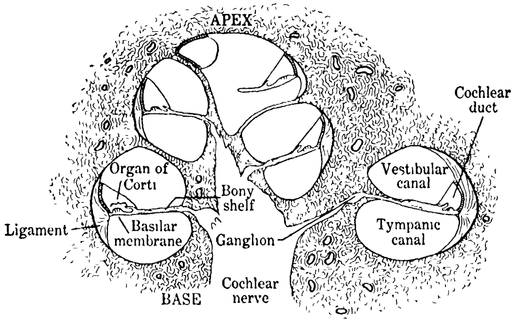


FIG. 30. THE COCHLEA IN MID-SECTION.
The section passes through the axis of the spiral, and cuts across the canals in five places.

base of the stirrup is attached. The second, the round window, which opens from the end of the tympanic canal, is covered only by a thin membrane.

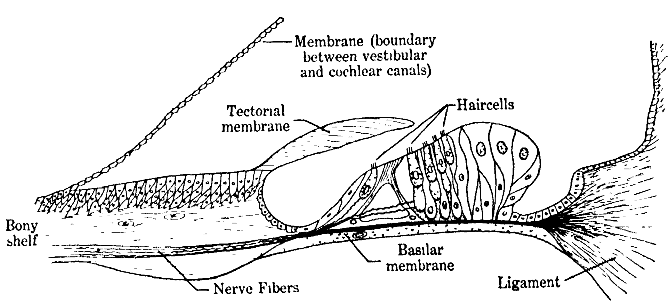


FIG. 31. SECTION THROUGH THE ORGAN OF CORTI.
See Fig. 30 for orientation of this section.

The *basilar membrane* lies between the vestibular canal and cochlear duct on the one hand and the tympanic canal on the other, and is attached to a bony shelf on the inner wall and to a ligament on the outer wall.

Upon the basilar membrane and within the cochlear duct lies the *organ of Corti*, a complex structure bearing hair cells, which are the auditory receptor cells. These cells contain the dendrites of nerve fibers which run inward through tunnels within the bony core around which the cochlear canals are wound, and finally come together to form the cochlear division of the eighth cranial nerve. A fold of fibrous tissue, the *tectorial membrane*, extends from the bony shelf into the cochlear duct and over the ends of the hair cells.

When the stirrup is momentarily forced inward in its response to a sound wave, it exerts pressure upon the perilymph of the vestibular canal. As the perilymph is practically incompressible and the cavities in which it is contained are enclosed by bone except at the oval and round windows, the pressure exerted at the oval window can be relieved only at the round window by bulging its membrane outward into the cavity of the middle ear. The pressure is communicated from the fluid of the vestibular canal to that of the tympanic canal through the basilar membrane, depressing this membrane in the process. When, an instant later, the stirrup is pulled outward under the influence of the sound, the fluid pressures are reversed, the basilar membrane is pulled upward, and the membrane of the round window is pushed inward. As the basilar membrane moves, the hair cells of the organ of Corti are stimulated, and they in turn excite impulses in the nerve fibers by which they are supplied.

THE PHENOMENA OF HEARING

Auditory experiences present a number of fundamental characteristics or attributes. Of these the most prominent are *pitch*, *loudness* and *complexity*.

Pitch. Tones vary continuously in pitch from low to high, or deep to shrill, as illustrated by the rising and falling note of the siren. In this instrument the sound is made by directing a jet of air against a revolving disk which bears around its periphery a row of holes. As the disk moves, the holes permit

the escape of air in separate puffs, the rate of which depends upon the number of holes and the rate of revolution of the disk. As the disk moves faster and the puffs increase in rate the pitch becomes higher. Thus it is shown that pitch varies with the frequency of sound waves, that slow frequencies give rise to low-pitched sounds and rapid frequencies to high-pitched sounds.

But sound frequency is not the sole determinant of pitch. It varies also, to a limited extent, with intensity. Tones of high frequency are raised in pitch when the intensity is increased, whereas tones of low frequency are lowered in pitch. For example, a tone of 5000~ at high intensity may seem equal in pitch to a weaker tone of 5010~, whereas a tone of 400~ at high intensity will seem equal to a weaker tone of 390~. The intermediate frequency at which a change of intensity has no effect upon the pitch is 3200~.

Finally, pitch is a function of sound composition. Tones have definite pitch; noises are essentially lacking in this characteristic. Many sounds, of course, are intermediate between tones and noises, and possess qualities of both these classes. Such sounds as the roll of thunder, the howling of the wind and the hiss of a steam jet present both tonal and noisy features: they are somewhat harsh and irregular, yet it is easy to recognize a characteristic pitch difference among them.

Loudness. A characteristic of tones and noises alike is loudness. It varies from the faintest whisper to the deafening thunder-clap, and depends upon both the energy and the frequency of sound waves. For a tone of a given frequency, the loudness varies with the energy in a fairly regular manner. But if sound energy is kept constant while frequency is varied, the tones will not appear equally loud, for the sensitivity of the ear varies greatly with frequency, according to functional relations that closely resemble the curve for the lower intensitive threshold in Fig. 34.

Since the sensitivity of the ear is a function of frequency, it follows that the loudness of a complex of several tones or of

noises does not depend in any simple way upon the total energy in the waves, but varies with the manner in which the energy is distributed among the various frequencies present.

Complexity. In respect to the attribute of complexity, sounds vary all the way from the simplest, clearest tone to the most bewildering noise. Simple or pure tones are rare in ordinary experience, for even the tuning-fork gives a complex tone unless care is taken to strike it gently. Musical instruments produce tones of various degrees of complexity, ranging from the simple note of the flute to the harsh jangle of the cymbals. Noises are often complex to the point of confusion, and typically are not steady, but continually changing in loudness and composition.

An examination of the sound waves which produce the above series of experiences shows that the pure tone arises from a sound wave of simple sine form, and as the wave form departs more and more from this elementary type the experience becomes increasingly complex. In Fig. 27 were shown the waves typical of the flute, the clarinet, the voice and the explosion of a sky-rocket.

Volume. Volume is a spatial attribute of sounds; it varies with frequency and intensity, and possibly with composition also. Low tones seem large and bulky; high tones seem small and thin. The relation to intensity is also a direct one: loud sounds seem large as compared with faint sounds.

Vocality. Sounds are said to have vocality in so far as they resemble the sounds of the voice, especially vowel sounds. Thus a low-pitched sound may resemble the sound *oo* as in 'moot,' and a high-pitched sound may resemble *ee* as in 'meet.' If we pronounce in rapid succession the vowel sounds *ōō, ō, ä, ā, ē* (as in moot, mote, mart, mate, meet), the concomitant rise in the pitch is easily recognized.

Tonality. Certain musical tones which are of different pitch have a similarity which is called tonality. If we run slowly up the tones of the piano, we find marked resemblances be-

tween certain notes. Most prominent are octave similarities: pairs of notes such as *c* and *c′*, which have frequency ratios of 1:2, are much alike in quality.

ANALYSIS

One of the fundamental laws of physical acoustics is that any complex periodic wave may be regarded as the sum of a series of simple harmonic motions of suitable amplitudes and phase relations, whose frequencies are in the ratios of the simple numbers 1, 2, 3, 4, 5, etc. This is Fourier's law.

The first member of the series, bearing the ratio number 1, is called the *fundamental*; the remaining members are *overtones*. The first overtone, with the ratio number 2, has a frequency twice that of the fundamental; the second overtone, with the ratio number 3, has a frequency three times that of the fundamental, and so on. Thus, for example, a complex sound may be considered as made up of a fundamental frequency of 100~, and overtones of 200~, 300~, 400~, and the like.

Any member of the series of frequencies may vary in amplitude or in phase, or may be absent (zero amplitude). The resolution of a complex vibration into the series stated by Fourier's law, with a determination of the component frequencies, amplitudes and phases, may be accomplished by mathematical methods, or, less laboriously, by a mechanical apparatus called a harmonic analyzer.

The ear is capable of analysis of the same general type as that performed by mathematical or mechanical methods. On listening to a complex tone we are able, especially after training, to 'hear out' the fundamental and several of its overtones. The fact that by ear we are able to resolve a complex tone into components that conform approximately to Fourier's series is known as *Ohm's law* of hearing.

In the sounds produced by many musical instruments, the largest amount of the total energy is in the fundamental, and

most of the remainder in the first few overtones. The form of distribution of the energy is usually fairly definite for a given instrument, as it depends in a measure upon distinctive mechanical features. It varies with the method of producing the vibrations. For example, distinctive forms are found in the piano, where strings are struck, and in the clarinet, where air is blown across a reed. It varies with the type of resonator, and is different, for example, in the piano, which has a sounding-board, and in the trombone, which has a variable tube of particular form. Since mechanical features vary somewhat among instruments of given design, owing to differences of material and workmanship, the distribution of energy among the components of the sounds will vary for particular instruments. Thus we have the difference between a good instrument and a poor one. The distribution will also vary with the method of playing, and thus we have one of the marks of distinction between the master and the mediocre performer.

The particular distribution of energy that is regarded as acceptable for an instrument or a player is determined by musical conventions, and the standards have varied during the course of musical development. During comparisons of instruments or performers, the fundamental note is usually kept constant, so that any differences that appear can be ascribed to the overtones and their relative amplitudes and phases. The subjective differences which arise from variations of the overtones under these conditions are called differences in *timbre*. Thus we speak of the difference in timbre between a clarinet and a violin, or between one voice and another. We can readily recognize a familiar instrument or singer by the timbre of the sounds produced.

The approximate character of auditory analysis is due to the peculiarities of the ear and its nervous connections, among which are the limited range of sensitivity, the different sensitivity to various sound frequencies and the distortion to which sound vibrations are subject in the middle ear and elsewhere.

SENSITIVITY OF THE EAR

The sound frequencies that can be perceived by the ear lie within a lower limit of about 20~ per second and an upper limit of about 20,000~. These limits vary in different individuals and may be greatly modified by disease. The upper limit varies with age; in young persons it may be a little above the value stated, whereas with persons above 25 years it usually decreases progressively, as shown in Fig. 32.

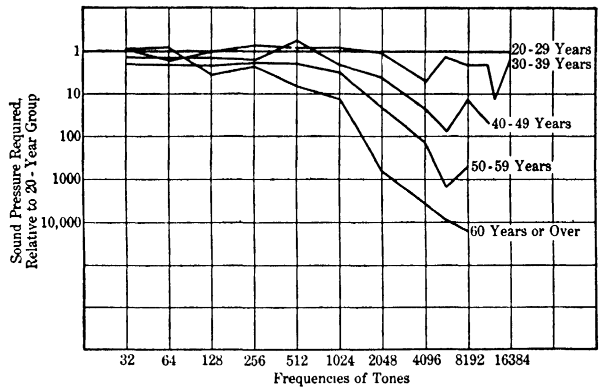


FIG. 32. RELATION BETWEEN AGE AND AUDITORY ACUITY.

Five age groups were examined under similar conditions. The curves show the averaged results, with the youngest (20-29 year) group as the base or 'normal' line For tones above about 512~ the older groups require a sound pressure many times larger than the 'normal,' and they fail completely to hear the extremely high tones. Data from Bunch, C. C., *Arch. of Otolaryngol.*, 1929, 6, 625-636.

Within the audible range of frequencies there is considerable variation in the ability to distinguish one pitch from another. For the middle region of tones, from about 500~ to 4000~, the capacity of appreciating differences of pitch can be stated as a nearly constant fraction of the frequency. A given tone within this region must be changed by an amount that for most persons is 3/10 of 1 per cent before a difference in pitch is noticed. For higher tones the fractional change must be slightly larger; for low tones they must be considerably larger, up to 4 per cent or more. These facts are shown in Fig. 33.

The ear also varies in its intensitive sensitivity, both as regards the limits of intensity to which it will respond and the changes of intensity that can be appreciated. If tests are made to determine for a great number of frequencies the least energy of sound that can be perceived, the results will give an audiometric curve such as the lower curve of Fig. 34. This curve represents the average of several individuals, and is smoother than

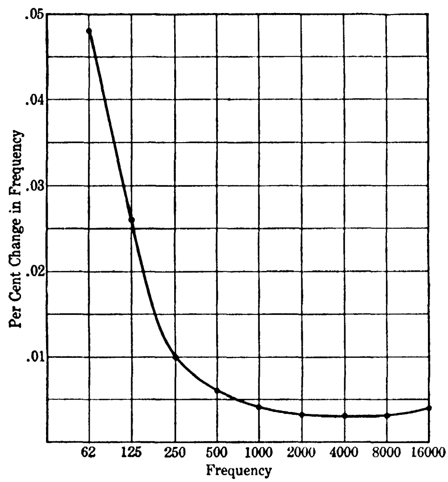


FIG. 33. PITCH DISCRIMINATION.

At low frequencies, the ability to distinguish pitch is considerably poorer than at frequencies above 1000~. Each tone is 40 decibels above threshold. Data from Shower, E. G. and Biddulph, R., *J. Acoust. Soc.*, 1931, 3, p. 284.

the curve for a single ear. It is evident that the ear is most sensitive in the middle range of 800~ to 4000~, and that its sensitivity drops off rapidly for higher, and somewhat more slowly for lower frequencies.

Not only is there a lower limit of sensitivity of the ear (lower intensitive threshold), but for practical purposes there is an upper limit also, because, as the intensity of any sound is raised, a point is finally reached where there is stimulation of pain receptors located in the middle ear or elsewhere. For low frequencies, up to about 1000~, pressure and kinesthetic endings

also are excited. The intensities where pain and other qualities enter may be taken, for practical purposes, as the upper intensitive threshold, for though the ear doubtless will operate above these intensities, the true limits cannot be investigated without risk of damage. The upper curve of Fig. 34 shows the general location of the upper intensitive threshold. Between this curve and the lower curve of the same figure we have the useful range of auditory experiences. It is evident that the range of

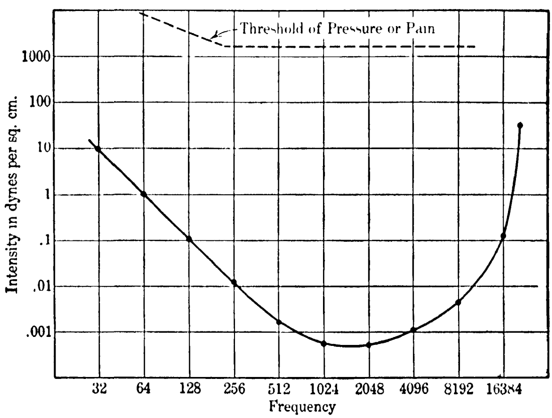


FIG. 34. AUDIOMETER CURVES, SHOWING THE SENSITIVITY OF THE NORMAL EAR. The lower curve represents the lower intensitive threshold; the upper curve indicates the upper intensitive threshold. The intensity is shown as the pressure exerted upon the eardrum, measured in dynes per square centimeter. After Wegel, R. L., *Ann. of Otol.*, 1932, 41, p. 770.

intensities is particularly great for the middle frequencies. For a 1500~ tone, the variation of pressures from bare audibility to the maximum that may safely be applied to the ear is more than a million-fold. And since the energy of sound waves is proportional to the square of the pressure, this variation in energy units is over a trillion-fold.

The ability to distinguish differences of intensity varies both with the frequency of the tones and with their level of intensity. Sensitivity is poor for faint tones and for tones near the limits of the frequency scale. Loud tones of the middle range have a relative difference limen of nearly constant value; such

sounds must undergo a pressure change of 9 per cent to produce a noticeable alteration in loudness. Fig. 71 on page 201 shows how the relative difference limen for intensity varies with intensitive level.

The essential constancy, within the limits shown, of the relative difference limens for pitch and intensity makes it convenient when dealing with either of these characteristics to use logarithmic rather than linear scales of measurement. Logarithmic scales for frequency were used in Figs. 32, 33 and 34, in which equal steps represent successive octaves. Another logarithmic scale for frequency is found in the tempered scale of modern music, where successive notes, representing semitone steps or intervals, differ in frequency by the constant ratio of 1:1.06 ($1:\sqrt[12]{2}$).

Logarithmic scales of intensity have also been used in Figs. 32 and 34. In both figures a given step represents a ten-fold change of pressure. A logarithmic scale of intensity in common use is the *decibel* scale (see p. 191), in which each successive unit represents an energy increase of 26 per cent over that preceding. It should be noted that this scale expresses a relation between quantities, and not their absolute values. Thus, to say that one sound is 20 decibels greater in intensity than another does not state their energies, but implies merely the ratio of these energies.

We may of course select a standard energy, and reckon the intensity in decibels above or below this as a base. A useful base is the threshold energy of the sound under consideration. Thus we may say that, from distances of 6 feet or so, the loudness of the rustle of leaves is about 10 decibels, of average speech is from 40 to 60 decibels and of a loud shout is from 80 to 90 decibels, all measured from the threshold of audibility.

We may also measure in decibels the amount that a sound is changed in loudness by various disturbing noises, using as a base in this instance the loudness of the sound under conditions of quiet. Thus it is found that the noises of the average business office have a disturbing effect on speech of 30 decibels. In a noisy office the disturbing effect may amount to 40 decibels,

in the New York City subway it is 60 decibels and in a boiler factory it may be as great as 80 decibels. By comparing these values with those given above for speech, it will be seen that in a noisy office understanding will be considerably reduced; in the subway and boiler factory ordinary speech will be inaudible, and even a loud shout will be perceived so faintly as to be largely incomprehensible (4).

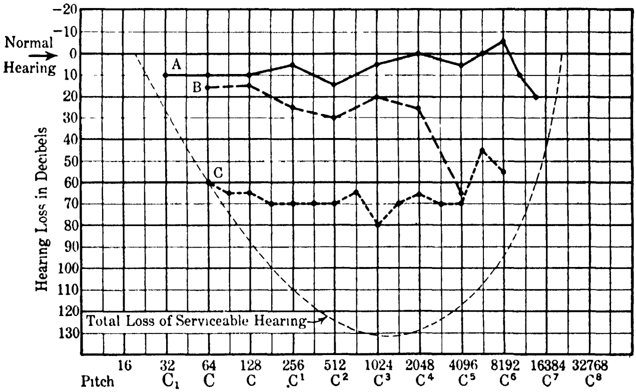


FIG. 35. TYPES OF AUDIOMETER CURVES.

A, practically normal hearing, except that the upper limit is slightly curtailed. *B*, a case of high-tone deafness *C*, a case of general deafness of serious degree. Data from Guild, S R, Crowe, S. J., Bunch, C. C, and Polvogt, L. M., *Acta Otolaryngol.*, 1931, 15, 292-301.

In many persons, as a result of diseases affecting essential portions of the auditory mechanism or its nervous connections, sensitivity is impaired in greater or less degree. The impairment takes many forms. In some persons, when the cochlea or auditory nerve is damaged, the deafness is total. In others, as shown in curve *C* of Fig. 35, the ear still functions, but all tones to be heard must be of considerable strength, far above that required by the normal ear. In many instances sensitivity is normal or nearly so for some frequencies, but reduced or lacking for others, as in curve *B* of Fig. 35, which illustrates high-tone deafness. Occasionally persons are found whose sensitivity is normal or practically so for most tones, but who have a

tonal gap, a restricted region of frequencies within which tones are heard only if their intensities are raised to an extraordinary height. Such a case is shown in *D* of Fig. 36. In other instances, essentially the converse of that just mentioned, all hearing is lost except for a limited range of frequencies where it is normal or at least fairly good; this remnant of hearing is called a tonal island (curve *E* of Fig. 36).

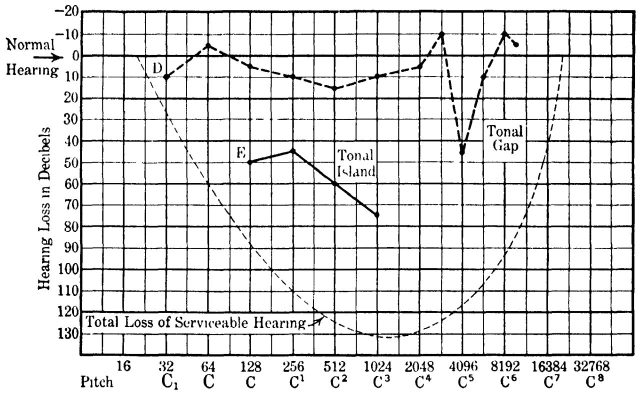


Fig. 36. Types of Audiometer Curves.

D, a case of nearly normal hearing, except for a partial gap in the region of 4096~. *E*, a tonal island extending from 128~ to 1024~ Data from same source as those of Fig 35.

Many persons, who are deaf to sounds conducted to the ear in the usual way, by the passage of aerial vibrations through the external auditory canal, are nevertheless able to hear when the sounds are applied directly to the bones of the head. Thus, often a person who cannot perceive the vibrations of a tuning-fork as it is held near the ear may hear them when the foot of the fork is pressed against the side of the skull or held between the teeth. The vibrations are then conducted directly through the bones of the skull to the inner ear, and thus may circumvent obstructions in the auditory canal or middle ear. When deafness exists for air-conducted but not for bone-conducted sounds, it is a sign that the essential tissues of the inner ear are functioning and that the defect lies in the conductive system of

outer and middle ears. When sensitivity is low for air and bone conduction alike, it is probable that the defect is of the inner ear itself or its nervous connections.

When deafness is only partial, it often may be compensated for by wearing a mechanical device that amplifies the sound waves before they are presented to the ear. One such device is the ear trumpet; another is a kind of portable telephone system. For persons with good bone conduction, best results are obtained from an instrument which applies the vibrations to the mastoid bone just behind the ear.

THE NATURE OF MUSICAL SOUNDS

Music consists of the use of sounds in complex patterns for esthetic expression. The patterns are formed by variations in *pitch, loudness* and *timbre*, according to accepted standards of rhythm, melody and harmony. These standards are not fixed, even in a given musical tradition; and they vary greatly for different peoples and in different ages.

In some very primitive music, *rhythm* almost exclusively appears as the form of expression. The sounds, which usually are noises from drums or other instruments of percussion, are woven into intricate patterns of a temporal and intensitive kind. This element of rhythm is found in all music, but in our own it is used in a relatively simple way owing to its subordination to melody and harmony.

Melody, which is the orderly succession or 'movement' of pitches in time, is the basic form in our own musical tradition. It probably arose as an accompaniment of the voice in poetic expression, and it still is the medium of song. Note follows note with changes of pitch upward or downward, usually in a well-defined rhythmic form. In a number of respects, melody resembles poetry and speech. It consists of a sequence of sounds arranged in phrases and measures, and it expresses a unitary theme. The constituent tones, like the syllables of speech, have little significance in themselves, but through their relations with others in the sequence they form a meaningful pattern.

Harmony, which is the use of many sounds in simultaneous combination, is a more recent acquisition. It sprang from the discovery that pleasing variety is achieved when a number of voices and instruments use different pitches together instead of singing or playing in unison.

The use of pitch in musical patterns led to the development of a great many instruments of various ranges and timbres. Some of these, like the trombone and violin, are capable of producing any pitch within their scope; but most, like the organ, piano, and horns, produce only certain fixed pitches. This latter limitation contributed to one of the outstanding characteristics of our music, the use of only a limited number of tones chosen from the audible range. These particular tones are represented by the notes of a musical scale.

It must be pointed out that the notes of a musical scale represent no absolute frequencies of sound, but rather a set of tones whose frequency ratios are fixed. For so long as the frequency ratios remain constant the intervals between the notes are sensibly the same, even though the absolute values of the frequencies vary over a considerable range. In other words, music is concerned fundamentally with relative rather than absolute pitch. It is a familiar fact that a melody may be sung or played at different regions of the scale and still is recognized as the same. The melodic relations, and also the harmonic characters, of a piece of music are not altered by changes of scale position.

Relative pitch ability, or the capacity to appreciate the intervals between notes when they are heard or when they are presented visually in musical notation, is common to persons of musical accomplishment. It is readily acquired through practice and may reach limits of high accuracy. Singing from memory or from printed music naturally depends upon this ability. But *absolute pitch* ability, or the capacity to identify a given isolated note either by stating its frequency or its musical name in reference to a standard system of tuning, is rare even among practiced musicians, and when it does exist it is relatively inaccurate. Whereas persons with musical training can judge relative intervals with an accuracy better than one-

fifth of a semitone, most persons even after special practice are unable to make judgments of absolute pitch within less than about 3 semitones (8). A few especially gifted individuals have greater absolute pitch ability than that just described, but usually their accomplishment is limited to a given instrument with which they are familiar, and thus probably depends upon the peculiarities of the several notes as regards timbre and loudness as well as pitch.

BEATS, COMBINATION TONES AND MASKING

Beats. When two tones are conducted simultaneously to the same ear, one of three phenomena may result. If the two tones differ but little in frequency, and are somewhere near the same intensity, there occur peculiar variations in the loudness of the sound, known as *beats.* If the frequency difference of the tones is increased, with tones both of which are fairly loud, a third tone, called a *combination tone*, is heard in addition to the two primary tones. But if one of the tones is considerably stronger than the other, it masks or obscures the other tone, and we hear only the stronger.

If two tones are identical in frequency there are no beats, and the combination of the tones differs from either of the tones alone only in loudness. The loudness in a given case depends upon the phase relation of the tones, and may vary anywhere from a maximum which equals the sum of the two intensities when the tones are in phase, to a minimum which represents the difference of the two intensities when the tones are in contrary phase. If the two tones are of equal intensity, the intensity at phase agreement is doubled, giving a loudness 3 decibels above that of either tone, and at phase opposition the two waves cancel and the result is silence. Intermediate phase relations give intermediate intensities.

But if the tones differ slightly in frequency, the phase relations change periodically, and beats occur at a rate that is directly a function of the frequency difference. If the frequency difference is 1, as with the tones 1000~ and 1001~,

there must be during 1 second a change from phase agreement to contrary phase and back again to agreement, which means that the intensity changes from maximum to minimum and back again to maximum. If the frequency difference is greater, then this variation of intensity must occur correspondingly more often during each second. The combination of two tones of adjacent frequency gives a wave of varying amplitude, illustrated in Fig. 37.

As the frequency difference of the tones is increased, the experience changes from a slow waxing and waning of the sound to a rather unpleasant 'pounding' in which changes of intensity within each beat are no longer perceptible, but each beat is a distinct impulse. With further increase of the

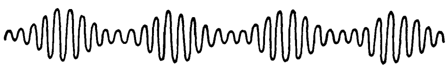


FIG. 37. BEATS.
Two tones of adjacent frequencies combine to give a curve of periodically changing amplitude

frequency difference the experience first passes into a stage of prominent roughness, and then fades and finally disappears. The point of disappearance of the beats varies with the region of the scale from which the primary tones are chosen; with high tones, beats as fast as 250 per second can be heard, but with low tones the limit is lower.

The appearance of beats between two tones of adjacent frequencies reveals a limitation of the ear's capacity of analysis. If analysis were perfect we should hear the two tones separately, with no beats. As a matter of fact, when the frequencies are close together we hear but a single tone of varying loudness. This evidently is not analysis, but a kind of synthesis. When the frequencies are farther apart, the two primary tones can be heard, and along with them a third tone which is beating. The third tone is called the *intertone*, and it has a pitch intermediate between the two primaries. The condition here is one of partial analysis. When the two tones are still

farther apart we get complete analysis: the tones are heard separately, with no beats.

The limited kind of analysis exhibited by the ear can be explained if we assume a particular distribution of action in the cochlea, in which different frequencies of sound, in the discharge of pressures from vestibular to tympanic canals, cause movements of different segments of the basilar membrane. If the mechanical properties of the cochlear system are continuously graded, then as frequencies are varied from low to high the segments of basilar membrane involved will vary regularly from one end of the cochlea to the other. Consequently, adjacent frequencies will involve adjacent segments. And if it is further assumed that the segment for a given frequency is fairly broad, then between frequencies there will be an amount of overlapping that is directly dependent upon their nearness in frequency.

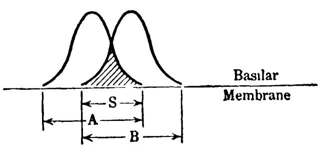


FIG. 38. THE ORIGIN OF BEATS. Two sound frequencies, 1000~ and 1010~, are represented as affecting adjoining regions of the basilar membrane, *A* and *B*. A common segment, *S*, is involved in both movements.

We can suppose, for example, that a tone of 1000~ produces up-and-down movements of the segment of basilar membrane *A* of Fig. 38, whereas a tone of 1010~ produces movements of segment *B*. A small segment, *S*, is involved in both movements. The amplitude of movement of this common segment will vary periodically, for when *A* and *B* are moving in the same direction, that is to say, are in phase agreement, the displacement will be a maximum, and when they are moving in opposite directions, or are in contrary phase, the displacement will be a minimum. The changes from maximum to minimum displacements for this common segment will occur at a rate that is equal to the difference in frequency of the two tones, for just so often will the phase changes occur. It will be seen further that the more separated the frequencies are, the less will be the overlapping, and the more prominent the primary regions *A* and *B*. With nearly complete overlapping, *A* and *B* lose their identity, and the pri-

mary tones are not heard. On the other hand, with slight overlapping, *A* and *B* are analyzed out, and the beating intertone fades into insignificance. The basic feature of this theory, which is the orderly distribution of the action of different sound frequencies along the basilar membrane, will be developed more fully in the last section of this chapter.

Combination tones. When the two tones that are led to the ear are fairly strong, and have a frequency difference of about 50 or above, one or more tones may be heard in addition to the primary pair. The most prominent of such tones is a difference tone, which has a frequency equal to the difference in frequency of the primaries. Under good conditions, when the primaries are loud, there can also be heard a summation tone, whose frequency is the sum of the frequencies of the primary tones. Thus, for example, the tones 700~ and 1200~ may produce a difference tone of 500~ and a summation of 1900~. If the primary tones contain strong overtones, there may be other combination tones, but these usually are so faint as to escape notice.

The appearance of combination tones is a further indication that the ear as a whole is not a perfect analyzing mechanism. New frequencies appear subjectively which would not be revealed in a Fourier analysis of the sound waves entering the ear. The most reasonable explanation of this phenomenon is that the new frequencies arise through distortion in the middle ear or elsewhere.

Masking. If the two tones that are led to the same ear differ considerably in intensity, the weaker is *masked* and only the stronger is heard. Masking is greater for tones close together in frequency, unless the tones are so near as to lose their identity in the production of strong beats. For ordinary tones which contain overtones the effect is greater for frequencies above the masking tone than for frequencies below it. The masking effect of a 1200~ tone is shown in Fig. 39. The curves show how high above their normal thresholds the various neighboring tones must be raised in order to be

perceived in the presence of a masking tone, at three different intensities, 40, 60 and 80 decibels.

An explanation of masking involves the same kind of assumption as was made to account for beats, namely, that sound frequencies are distributed regularly in their action on

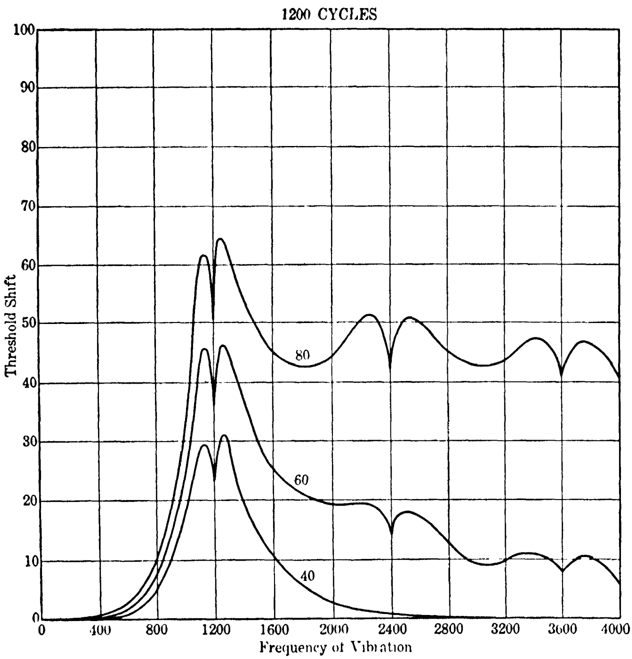


FIG. 39. AUDITORY MASKING.

Three curves are given to show the masking effects of a 1200~ tone at intensities of 40, 60, and 80 decibels above threshold. Modified from Wegel, R. L., and Lane, C. E., *Phys. Rev.*, 1924, 23, 266-285.

the basilar membrane. If the regions corresponding to two tones overlap considerably, and one tone is much the stronger, the stronger tone will stimulate most of the nerve endings in that region and will leave few or none available for the other frequency. The second is then masked or obscured by the stronger tone.

AUDITORY LOCALIZATION

The cues that are involved in the localization of sounds may be divided into two classes: primary or binaural cues, which depend upon differences of stimulation of the two ears; and secondary cues, which are not significantly different for the two ears.

A difference in the stimulation of the two ears, if within the proper limits, results in the perception of a single sound displaced to the right or left of the median plane. For simple tones, this difference in stimulation may be one of *intensity* or of *phase*, or a combination of the two. For complex tones and noises, the difference may be of *intensity, priority, complexity* or possibly of *phase*, or a *combination of these.*

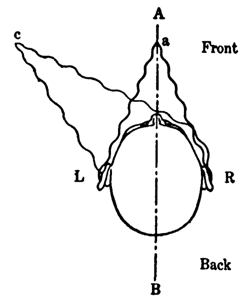


FIG. 40. THE LOCALIZATION OF SOUNDS.

A sound at *a* will stimulate left and right ears equally But a sound at *c* will stimulate the left ear sooner, more strongly, and in different phase.

The direction of the displacement of the sound depends upon which ear is favored by the binaural difference in stimulation, and the amount of displacement depends upon the magnitude of this difference.

If the source of sound is in the median plane, as at point *a* in Fig. 40, or anywhere on the surface that may be drawn through *AB* perpendicular to the paper, then it is equidistant from the two ears, and, barring any intervening object, must stimulate the two ears in precisely the same manner. But if the source is moved to point *c* it will stimulate the two ears differently. Because it is nearer to *L*, it will stimulate that ear sooner. For the same reason, and also because *R* is somewhat in a 'shadow' cast by the head, it will stimulate *L* more strongly. Further, because of the difference in distances *cL* and *cR*, there will be a difference in phase, except for those few tones for which the difference in distance equals some integral number of wave-

lengths. The phase difference will vary for different tones, because it depends upon the relation between wave-length and the difference in distance.

Under normal conditions of stimulation the above cues operate concurrently. But under experimental conditions they may be controlled individually and separately studied. The best methods for such control involve the use of electrical apparatus. Separate telephone receivers are placed over the ears, and to these are conducted oscillating electrical currents whose magnitude, phase and time of incidence may be changed at will.

If time and phase are kept constant, and the sounds at the two ears are varied in intensity, it is found that a shift of the apparent source of sound from the median plane begins to appear when the intensity in one ear is about 4 decibels greater than that at the other. As the ratio of intensity is further increased the sound seems to move toward the side of greater intensity, until eventually it appears to be directly opposite the favored ear.

If the intensity and time are kept constant, the localization of tones varies with the phase. The sound is localized in the median plane so long as the tones at the two ears agree in phase. But if the phase is made to lead in one ear, the sound seems to be shifted toward that side. The lateral shift becomes maximal for a phase difference of nearly 180°. At exactly 180° the phase is equivalent for the two ears, and localization is median or uncertain. A phase lead of more than 180° is the same as a lagging phase, and localization shifts to the opposite side.

The change of localization with phase is most easily illustrated by applying a given tone to one ear, and to the other ear a tone of the same intensity but of slightly different frequency. The difference in frequency may be so small as to be imperceptible, but the physical effect is to give a periodic variation of phase relations. The situation is one that would give beats if both tones were led to the same ear. But here, with the tones in separate ears, beats do not occur. The fused sound

maintains a steady intensity, but it seems to shift back and forth between right and left sides of the observer. This binaural shift from right to left and back again occurs at a rate that is numerically equal to the difference in frequency of the tones.

For simple tones, time differences are inherently related to phase differences. But for noises, priority may be considered independently. If intensity and phase are kept constant, and the binaural sounds are varied in time of incidence, a shift of localization begins to appear when the priority at one ear exceeds 30 microseconds (30 millionths of 1 second). As the time difference increases, the apparent source moves farther from the median plane toward the ear where the stimulation is prior, and finally is directly opposite that ear for a time difference of 630 microseconds. There is no change of localization for further increases of priority up to 2000 microseconds, but for this and greater values the two sounds no longer fuse into one, but are heard separately, one on one side and then one on the other.

It is known that complex sounds are much easier to localize than simple tones. This may be due to a difference in the complexity of the sounds at the two ears. It happens that sounds of low frequency, with their relatively great wave-lengths, do not maintain straight paths after passing obstacles, but readily bend around them. Therefore the head does not give an appreciable sound shadow for low tones, and a source on the extreme left will stimulate the right ear nearly as strongly as the left. At 256~ the screening effect of the head is less than 0.5 decibel, and is therefore negligible. At 512~ and 1024~ the screening effect reaches 3 and 5 decibels respectively, which is enough to be significant, and for still higher tones it becomes increasingly important. Hence it follows that sounds which are made up of both low and high frequencies will not appear of the same degree of complexity at the two ears, if the source is out of the median plane. The ear toward the source will receive the waves in their true complex form, while the ear that is away from the source, and in the head's shadow, will receive the sound with the high frequencies considerably

weakened. This difference in complexity is probably appreciated in terms of localization.

Because there is little screening action of the head at low frequencies, differences of intensity are much less significant for low than for high tones. But a difference of phase, on the other hand, is particularly important for low tones. Its effect grows progressively less for high frequencies until it becomes of little value beyond about 1500~. For complex tones and noises which contain high frequencies, a difference in intensity operates in much the same way as for high tones. Phase differences, however, will usually be ambiguous as a cue to the localization of complex tones, since the phase relation varies for the different components of the wave. Time differences are very important in the localization of noises, but, as mentioned above, they are not distinct from phase differences for pure tones. Differences of complexity are naturally present only for complex tones.

It is evident that there are considerable differences in the cues available for the localization of various kinds of sound. For low tones, the binaural differences are chiefly of phase, and to a slight degree of intensity. For high tones the differences are chiefly of intensity. For complex tones and noises there may be differences of intensity, time and complexity. These differences in available cues are reflected in the relative ease of localizing different types of sound. High tones are localized only with difficulty, low tones somewhat more readily, and noises with comparative ease. A consideration of the relative difficulty of localizing a cricket's chirp, the whistle of a locomotive and the sound of a person's voice will show these differences.

It has been shown that the binaural cues are effective only in determining the angle of a source of sound to the right or left of the plane drawn between the two ears. These cues give no aid in determining the position up and down, or front and back, nor do they determine the distance from the observer. As shown in Fig. 41, all points that lie on the surface of a

regular cone drawn with its apex on the line joining the two ears are equivalent so far as primary localization is concerned.

However, our actual judgments of localization are more definite than mere angular direction to the right or left, since many additional, secondary cues enter into their formation. These cues are in part auditory, including the absolute loudness, the timbre and the temporal pattern of the sounds, and changes in the character of the sounds in correlation with movements of the head or body. In addition to these are extra-auditory cues, such as the perception of visual objects which may be considered as possible sources, and our knowledge of things in the environment which are reasonable sources of the sounds that we hear. These aspects of the problem of localization are considered further on pp. 236-238.

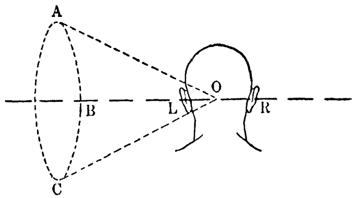


FIG. 41. THE LIMITS OF AUDITORY LOCALIZATION.

All points on the surface of the cone *ABCO* are practically equivalent in the relative stimulation of the two ears.

AUDITORY THEORY

In the second section of this chapter the discussion of the action of sound vibrations on the ear was extended to the stimulation of the hair cells of the organ of Corti. In this section we shall attempt to follow in further detail the process of hearing in the cochlea, and to inquire into the nature of the action set up in the fibers of the auditory nerve.

We have already found, in the discussion of nervous activity in general in Chapter 2, that the variations possible to a bundle of fibers like the auditory nerve are limited to two dimensions. These are the particular fibers activated by a given stimulus and the rates of impulses in these fibers. It is the particular task of a theory of hearing to show how the manifold nature of our world of sound can be expressed by these simple variables.

The features that are chiefly concerned in theories of hearing are pitch, loudness, analysis and localization. We have already discussed analysis in connection with the phenomenon of beats, and there have made the assumption that different parts of the cochlea, or, more specifically, of the basilar membrane, are involved in the action of different frequencies of sound.

This assumption was made by Helmholtz to provide an explanation of pitch. According to his theory, different transverse fibers of the basilar membrane vary in mechanical properties so as to give a progressive tuning from low to high, like that of the strings of the piano. This tuning would cause a certain region of the membrane to respond most actively to some particular frequency, the resonant frequency, just as in the piano certain strings will be set into *sympathetic vibration* if, with the damper raised, a given note is sung. The action of the basilar membrane at any given place would excite the particular nerve fibers that supply that region, and would thereby lead to the hearing of a tone of a given pitch.

Another hypothesis has been made for the explanation of pitch. This is called the 'frequency' hypothesis to distinguish it from the 'place' hypothesis mentioned above. It supposes that the frequency of sounds is represented in a direct manner in the frequency of impulses transmitted by the auditory nerve.

The above two hypotheses have traditionally been considered as rivals, but the theory to be presented here, which is known as the resonance-volley theory, incorporates the fundamental features of both.

Several lines of evidence lead us to accept the principle of the place hypothesis. The basilar membrane varies in width from base to apex, so that the fibers which extend transversely from the bony shelf on the one side to the ligament on the other are about three times as long near the apex as they are near the base. Moreover, the ligament is larger near the base, and presumably exerts more tension there than near the apex. Finally, the masses of cells lying on the membrane and weighting it are smaller near the base and larger toward the

apex. The mechanical differences are thought to give a continuous gradation of resonant properties from base to apex, so that the transmission of high frequencies from oval to round window involves the basal regions of the membrane, while the transmission of low frequencies involves the apical regions.

Further evidence regarding the place hypothesis comes from experiments in which animals are stimulated for a long time with a loud tone. Examination of the cochlea by histological methods after such stimulation has usually shown changes in the structure of the organ of Corti. High tones cause changes in the basal region, but the effects of low tones are uncertain.

The experiments just mentioned are restricted to animals, but evidence of a comparable kind is gradually being accumulated for human beings. This is done by a histological study of the inner ears of persons whose hearing has been tested and whose temporal bones have become available after their death. The results show a relationship between deafness and local degenerations within the cochlea. In a significant number of subjects with high-tone deafness, regions of degeneration are found in the basal part of the organ of Corti, and also in the nerve fibers supplying this region. When deafness is general, the degenerations are more widespread (3).

The above evidence is in agreement in indicating that the basal portions of the cochlea are involved in the hearing of high tones. The relationship between stimulus frequency and anatomical region is not highly specific, but a given tone may involve a fairly broad band of the basilar membrane, containing many transverse fibers. The evidence is not so clear for the action of low frequencies, but it is probable that they involve relatively large bands. The very low frequencies, indeed, may involve practically the whole extent of the membrane.

Our acceptance of the frequency hypothesis, in some form or other, is warranted by the following two lines of evidence. The phenomenon of phase localization seems to indicate that, for low tones at least, the phase relations of tones presented to the two ears are somehow preserved in the nervous action. The phase relations would be preserved most strictly, and at

the same time most simply, if the entire wave-form could be preserved in the nerve transmission process. But since, as we know, a nerve fiber acts with only discrete impulses, and cannot reproduce an undulating wave, this preservation of wave-form cannot occur in any simple manner. It is sufficient for the explanation of phase localization to suppose that there is some fairly definite relation between the number of waves in the stimulus and the number of nerve impulses.

Direct evidence of a correlation between the frequency of the sound and the frequency of nerve impulses is obtained from experiments on animals in which an electrode is placed on the auditory nerve or its projections in the central nervous system, and the nerve impulses studied during stimulation of the ear by sound. It is found that, for low and intermediate tones, impulses passing up the auditory nerve represent accurately the frequency of the sound waves. The upper limit of frequency for which this relation remains true has not yet been determined (11).

The above fact, that low and intermediate frequencies are represented in the action of the cochlear nerve, does not mean necessarily that individual fibers of the nerve operate at correspondingly high rates. Nerve fibers are limited in their actions, for after a given impulse they require a period of recovery, the refractory period, before they are again stimulable. The lengths of this period in mammalian nerves are probably too great to permit, in a given fiber, the continuous passage of impulses at a rate higher than a few hundreds per second (see p. 18f.).

A way around this limitation is provided by the *volley* hypothesis. This hypothesis states that a number of nerve fibers, each responding at a relatively slow rate, may cooperate to establish a more rapid rate, just as a drummer, by using two drumsticks, can beat a tattoo of double the rate of each stick. At very slow frequencies, each fiber may of course respond at every wave of the stimulus, but as the frequency becomes higher the fibers respond at every second wave, or every third, and so on. Different fibers, since their rates of recovery differ

slightly, and because the strength of stimulation varies over the region of the basilar membrane involved, become out of step with one another. That is to say, they come into the total discharge of impulses at different times, and if there are enough fibers there can be several impulses for each wave of the stimulus. Though many of the fibers are out of step with one another, they are all in synchronism with the waves of

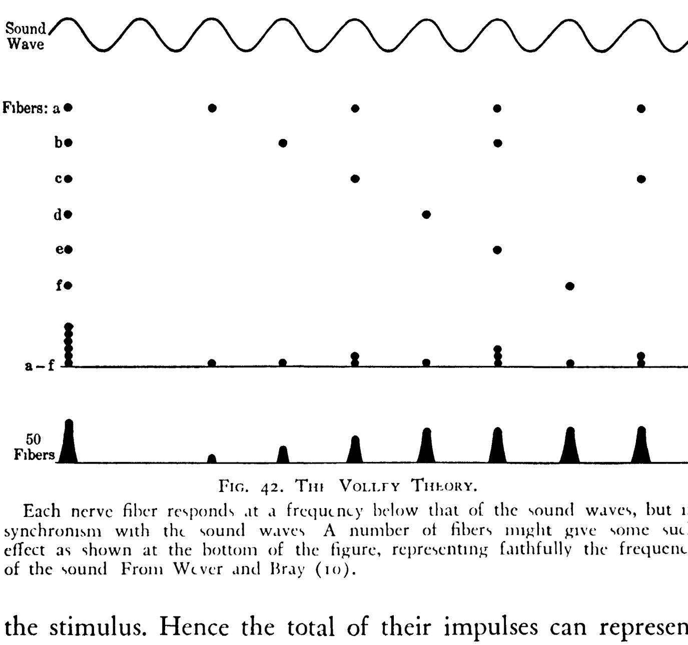


FIG. 42. THE VOLLEY THEORY.

Each nerve fiber responds at a frequency below that of the sound waves, but in synchronism with the sound waves. A number of fibers might give some such effect as shown at the bottom of the figure, representing faithfully the frequency of the sound. From Wever and Bray (10).

the stimulus. Hence the total of their impulses can represent with a high degree of faithfulness the frequency of the sound.

This hypothesis is illustrated in Fig. 42, where the upper curve represents the sound waves, and *a, b, c, d, e, f* represent six different nerve fibers, which fire at the times indicated by the dots, reading from left to right (10). The total impulses of these six fibers are next shown, and finally the result to be expected from fifty fibers. Note that the figure indicates but one stimulation of a nerve fiber during a given cycle of the stimulus, and this at some fairly definite phase position. The

time relation between the waves of stimulus and firing of the fibers is not assumed to be absolutely accurate, as shown by the scattering of the responses for fifty fibers into hill-shaped groups. The inaccuracies will not be serious for ordinary frequencies, but they might well become so for very high frequencies.

Loudness, on this theory, depends upon the total number of impulses passing through the auditory nerve per unit of time. As the energy of the stimulus is increased, the movements of the basilar membrane increase both in amplitude and in extent. The result is the stronger stimulation of the sensory endings in the region first concerned and the stimulation of more endings both in that region and on either side, owing to the greater spread of movement. In the individual nerve fibers the effect of stronger stimulation is to increase the rate of impulses, since with stronger stimulation a fiber can respond earlier in its refractory period. This change in the rate of firing of the fibers does not impair the synchronism with the sound waves. It simply means that a given fiber enters the volleys more often. The fiber may, for example, act at every second wave instead of at every third (see Fig. 68, p. 194).

This type of explanation of loudness is in accord with what we know of sensory intensity in general, as discussed in Chapter 8 (see pp. 193-196).

Most of the remaining phenomena of hearing have been discussed sufficiently to indicate their general relationships to this theory. The analysis of complex sounds is explained as based primarily upon the distribution of different components over the linear extent of the basilar membrane. And since this distribution is not highly specific, but adjacent frequencies overlap somewhat, we have beats and masking as products of their interaction. Marked changes of sensitivity, such as high-tone deafness and tonal gaps, are explainable as due to local disturbances of the mechanical system or its nervous connections. Finally, as already suggested, the localization of sounds in terms of phase can be accounted for on a basis of the fairly

accurate time relations between the waves of the stimulus and the impulses that are transmitted by the auditory nerve.

REFERENCES

1. BEATTY, R. T. *Hearing in man and animals.* London: G. Bell and Sons, 1932.
2. BORING, E. G. Auditory theory with special reference to intensity, volume, and localization. *Amer. J. Psychol.,* 1926, 37, 157-188.
3. CROWE, S. J., GUILD, S. R., and POLVOGT, L. M. Observations on the pathology of high-tone deafness. *Bull. Johns Hopkins Hospital,* 1934, 54, 315-379.
4. FLETCHER, HARVEY. *Speech and hearing.* New York: Van Nostrand Co., 1929.
5. M'KENDRICK, J. G., and GRAY, A. A. The ear. In E. A. Shafer's *Textbook of physiology.* Edinburgh: Y. J. Pentland, 1900. Vol. II.
6. MILLER, D. C. *The science of musical sounds.* New York: Macmillan Co., 1926.
7. OGDEN, R. M. *Hearing.* New York: Harcourt, Brace and Co., 1924.
8. WEDELL, C. H. The nature of the absolute judgment of pitch. *J. Exper. Psychol.,* 1934, 17, 485-503.
9. WEVER, E. G. The physiology of hearing: the nature of response in the cochlea. *Physiol. Rev.,* 1933, 13, 400-425.
10. WEVER, E. G., and BRAY, C. W. Present possibilities for auditory theory. *Psychol. Rev.,* 1930, 37, 365-380.
11. WEVER, E. G., and BRAY, C. W. The nature of acoustic response: the relation between sound frequency and frequency of impulses in the auditory nerve. *J. Exper. Psychol.,* 1930, 13, 373-387.
12. WILKINSON, G., and GRAY, A. A. *The mechanism of the cochlea.* New York: Macmillan Co., 1924.
13. WOOD, A. B. *A textbook of sound.* New York: Macmillan Co., 1932.

Chapter 6

TASTE AND SMELL

Sight, hearing, taste and smell are known as the four special senses in contradistinction to somesthesis which is general bodily sensibility. Taste and smell are thought to be allied because the qualities of each depend upon the chemical action of the stimulus, because they appear to be descended evolutionarily from the single chemical sense of fishes and because in civilized man their chief function is the common one of the perception and appreciation of food. There is some evidence that smell at least may provide as great a range of phenomena as sight or hearing, but our knowledge of both senses remains meager because we have never gained that degree of insight into the nature of their stimuli which we have for color and tone.

TASTE

Receptors. Taste is subserved by specialized spindle-shaped receptors, the *taste-cells*. These cells, which bear hair-like processes on their outer ends, are assembled in groups of from two to twelve to form *taste-buds* (Fig. 43). Each taste-bud has at its outer end a small opening, the *taste-pore*, through which stimulus-substances in solution can penetrate to affect the receptors.

Taste-buds are flask-shaped, about 0.06 mm. long and 0.04 mm. in diameter. They are scattered irregularly on the surfaces of the oral cavity: on the tongue, the epiglottis, the larynx and parts of the throat. Children have many more of these organs, especially on the tongue and throat, than do adults, and presumably therefore have more reason for crowding their mouths with pleasant-tasting food.

This chapter was written by Michael J. Zigler of Wellesley College.

In general the taste-buds are inlaid in the epithelium of the mucous membrane. On the tongue, however, they appear almost always in connection with the various kinds of papillae which are to be seen all over the surface of the tongue. There may be anywhere from three to twenty taste-buds in those smaller papillae which have taste-buds and which are scattered widely over most of the surface of the tongue, but the large 'vallate' papillae at the base of the tongue may have fifty or a hundred taste-buds each.

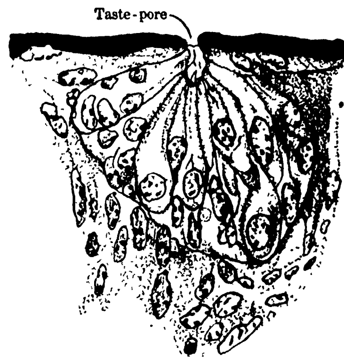


FIG. 43. DIAGRAMMATIC VIEW OF A TASTE-BUD, SHOWING A NUMBER OF TASTE-CELLS GROUPED TOGETHER TO FORM THE TASTE-BUD.

The taste-pore is the opening through which the sapid stimulus reaches the taste-cells. Taste-buds are associated with individual papillae on the tongue. After Kolmer.

Quality. *Sour, saline, bitter* and *sweet* are the four primary qualities of taste. There are also intermediate qualities, and the relationships among qualities have been represented by a tetrahedron, as in Fig. 44. The following are examples of sapid substances that give rise to qualities intermediate between the six pairs of primaries, *i.e.,* qualities that would be placed on each of the six edges of the tetrahedron:

Between saline and sour	carbonate of soda.
Between saline and bitter	potassium bromide.
Between saline and sweet	alkali.
Between sweet and bitter	acetone.
Between sweet and sour	acetate of lead.
Between bitter and sour	potassium sulphate.

There are also tastes that resemble three different primary qualities in varying degrees and even all four of the primaries. Such tastes would be represented as lying respectively in the surfaces and in the interior of the tetrahedron, which must therefore be thought of as a solid figure.

These four primaries and their intermediates do not furnish nearly so rich a variety of tastes as a common knowledge of gastronomic esthetics might lead one to expect. However, food owes its rich diversity, not only to taste, but also to touch and smell. A careful observer can often distinguish between the gustatory, olfactory and tactual components of a 'taste' complex, and anyone can demonstrate the role of smell in the flavor of food by the simple expedient of closing his nostrils. When smell is excluded, claret tastes like vinegar, and beef tea resembles a weak solution of salt. However, the beef tea may still be distinguished from a saline solution because its greasiness adds a tactual component that identifies it.

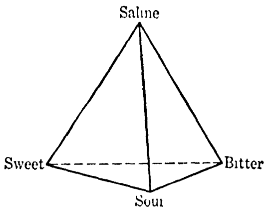


FIG. 44. TASTE TETRAHEDRON. A figure proposed by Henning to show the relationships of the various gustatory qualities to the four primary tastes. The figure must presumably be regarded as a solid.

Stimulation. The stimuli to taste are solutions of certain chemically effective substances that enter the taste-bud through the taste-pore. Typical stimuli for the four primary qualities are as follows:

For sour	hydrochloric acid.
For saline	sodium chloride (common salt).
For bitter	quinine.
For sweet	cane sugar.

In the preceding section we have noted typical stimuli for six intermediates between the primaries. A gross generalization, to which many exceptions can be found, is to say (*a*) that sour depends upon acids, acid-salts and acid-producing substances, (*b*) that saline is aroused by a large number of inorganic salts, especially the salts of chlorine, bromine and iodine, (*c*) that bitter is caused by alkaloids and (*d*) that sweet depends upon carbohydrates.

Attempts further to specify the nature of the chemical action have not been very successful. It seems certain that the essential stimulus to sour is the free hydrogen ions of acid substances

in solution. Saline appears to depend on the presence of the negative ions of chlorine, bromine and iodine, named here in order of their effectiveness. Solutions containing the positive ions of magnesium, ammonium and calcium are usually bitter, but this primary quality can also be aroused by many organic substances in which the structure of the molecule is more important than the elements which it contains. For the most part sweet is aroused by carbohydrates—the sugars and the alcohols —although there are some few inorganic compounds, like acetate of lead, which are sweet. It is plain that a really illuminating discovery about the chemical nature of gustatory stimulation has yet to be made.

Fusion, compensation and contrast. When the stimuli to different taste qualities are mixed, the resultant perception is a *fusion* in which the components can be distinguished, sometimes with ease and sometimes only with difficulty. Fusions at low intensities are least readily resolved by attention into their components. Mixtures that involve bitter are usually more easily analyzed than those that do not. The taste of sweet chocolate (sweet-bitter) is a fusion, but it is not so intimate a fusion as the taste of weak lemonade (sour-sweet) or of a 'marguerite' (a confection on a salt cracker, hence salt-sweet). Modern salads present a great variety of gustatory fusions. However, all these gastronomic delicacies involve touch, temperature and smell as well as taste. In fact, smell is esthetically more important than taste in the sensory quality of food. A pure taste, without olfactory or tactual components, is a relatively rare and unimportant experience in everyday life.

There is mutual *compensation* in the mixtures of stimuli to sour, saline and sweet, that is to say, the components in the fusion are perceived as less intense than each would be alone. Bitter appears not to enter into these compensatory relations. Compensation is greater for weak stimuli. Lemonade is less sour and also less sweet because the acid and sugar weaken each other's effects, and this mutual diminution is greatest

in weak lemonade. Compensation is never total, as it may be with complementary colors.

It is also true that the stimuli to sour, saline and salt, when applied separately to different parts of the tongue, tend toward mutual enhancement, a phenomenon of *contrast.* Thus a subliminal sweet solution, applied to one side of the tongue, may become perceptible if a moderately strong salt solution is applied to the other side of the tongue. The stimuli for bitter again seem not to enter into these relationships. However, contrast can have little gastronomic importance, since food does not ordinarily maintain spatial patterns in the mouth.

Adaptation. *Adaptation* to taste is slower than for smell or vision, but long exposure of the receptors to a strong stimulus diminishes and may even ultimately abolish sensitivity to one class of primary stimuli, leaving the other three classes unaffected. The phenomenon can be demonstrated more readily if a single papilla is selected for experimental exposure and test. Such a papilla, adapted to salt, may still remain sensitive to sweet, sour and bitter. Luncheon, immediately after a morning of experimentation in the laboratory with the mouth continuously filled with strong saline solution, is a sorry affair, quite flat and tasteless.

As with color and brightness, taste adaptation seems, in some cases at least, to lead to *successive contrast.* Everyone knows that a plum may be too sour after candy, but quite sweet after grapefruit.

Sensitivity. The intensitive threshold measures taste sensitivity, but such measures lack general significance because different substances are so very differently effective. For instance, Table VIII shows how different are the threshold concentrations for four different alcohols in the cases of both taste and smell. The table suggests that in this case sensitivity may vary directly with molecular weight, but this generalization must be considered with great caution.

Sensitivity varies slightly in different regions of the oral cavity. The general rule is that the tip of the tongue is most

sensitive to sugar, the back of the tongue to quinine sulphate and the edges of the tongue to hydrochloric acid. Sensitivity

TABLE VIII

Threshold Values for Taste and Smell

Minimal amounts of four alcohols in molar solution necessary for the detection of taste and of smell. A relation to molecular weight is suggested. From Von Skramlik, after Hallenberg.

Stimulus	Molecular weight	Taste threshold	Smell threshold
Methyl alcohol	32 0	1 62	0 025
Ethyl alcohol	46 1	0.45	0 0016
Propyl alcohol	60.1	0.16	0 0004
Butyl alcohol	74 1	0 07	0.00004

to salt is approximately the same in all regions of the tongue. The more exact data for this spatial variation are given in Table IX. This lingual topography is said to find its consequences in certain facial expressions. There is, for instance, the 'complete pout,' in which the lips are opened and protruded

TABLE IX

Threshold Values for Individual Papillae in Different Parts of the Taste Field

In terms of molar concentration. From Von Skramlik, after Kiesow.

Locality	Quinine sulphate (bitter)	Sodium chloride (saline)	Hydrochloric acid (sour)	Cane sugar (sweet)
Tip of tongue	0 00000039	0 043	0 0029	0 014
Edge of tongue (right)	00000027	041	002	022
Edge of tongue (left)	00000028	043	0017	021
Base of tongue	00000067	048	0045	023
Soft palate	000004	034	0041	.044
Arch of palate (right)	0000048	102	0027	044
Arch of palate (left)	0000054	085	0036	.058
Uvula	0000054	153	0055	.073
Inferior tongue (right)	000134	051	011	175
Inferior tongue (left)	000134	051	0137	146

and the throat becomes thickened because the base of the tongue is lowered. The resulting expression is one of distaste,

and functionally it is this attitude that an infant takes in getting an unpleasant bitter substance out of his mouth: the base of the tongue, most sensitive to bitter, is lowered so that it gets minimal stimulation, and the lips are arranged as in spitting something out of the mouth. There is also the 'sweet mouth,' that may perhaps have some relation to the smile. In it the lips are held tightly together and drawn slightly in, and the tip of the tongue, most sensitive to sweet, presses forward—altogether an attitude that would yield the maximum of the pleasant sweet stimulation.

Physiology. In general the individual papillae are sensitive to more than one kind of primary stimulus; many are sensitive to all four, but a few respond to only one. However, the fact that usually a number of taste-buds are associated with a single papilla prevents us from deciding as to whether a single taste-bud mediates several or only one of the primary taste qualities.

It has usually been assumed that there are four different systems of nerve fibers, one for each of the four primary qualities—sour, saline, bitter and sweet. However, there is no proof at present that qualitative differentiation in taste is based upon any such simple anatomical distinction.

SMELL

Receptors. The sense organ for smell consists of an aggregation of *olfactory cells* in the uppermost recess of the nasal cavity. These cells are the receptors. They are spindle-shaped with large nuclei, and their central ends are continuous with the afferent nerve fibers (Fig. 45). The region which contains these receptors is about 600 sq. mm. in area and is clearly delimited from the rest of the membranous lining of the nose. The main current of the breath does not reach the olfactory organ, which is stimulated only by eddy currents that arise from the principal stream. For this reason sniffing increases sensitivity greatly. The inaccessibility of the organ makes the great sensitivity of many persons to odors surprising (see

p. 152), and it also explains why catarrhal conditions, so common in civilized man, render many other persons relatively insensitive.

Quality. We have names for the sensory qualities of colors, tastes, tones and tactual impressions: *green, bitter, F-sharp, pain.* There is, however, no such vocabulary for smell. Odors are named only by the objects that give rise to them. Such objects as the following may mean either a thing or the odor of the thing: cinnamon, cloves, fish, rose, onion, camphor, sulphur, cologne, dog, stable, garden, wharf, bakery, orchard. Even such general classificatory names as *flowery, fruity, spicy, resinous* and *burned* are derived from classes of objects which have such odors.

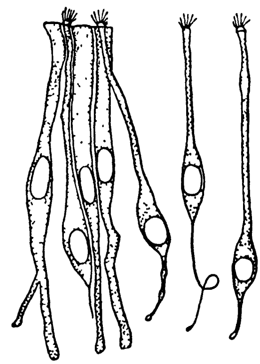


FIG. 45 OLFACTORY CELLS, TOGETHER WITH SOME SUPPORTING CELLS

From the olfactory regions of the nose From Parker, after Von Brunn

The world provides a great diversity of olfactory experiences, and there has never been an accurate classification of them or a perfectly correct system of relations among them. However, the triangular prism of Fig. 46 is a great convenience in the designation of odors. The figure establishes six primary classes of odors: *flowery, fruity, spicy, resinous, putrid* and *burned.* Odors of intermediate qualities lie on the edges or in the surfaces of the prism, being thus established in relation to two or three or four of the primary classes. Thus the smell of arbor vitae is flowery-fruity-spicy-resinous, and the odor of onion is mostly putrid, but also

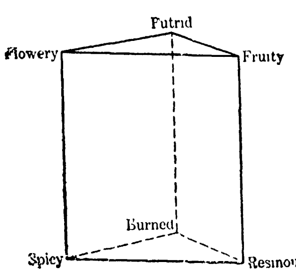


FIG. 46. SMELL PRISM.

Figure proposed by Henning to show the relationships among odors

somewhat flowery, burned and spicy. It is not known whether there are odors lying within the prism that have relation to all six of the primary classes. In general, putrid and burned odors tend to be unpleasant and the odors of the other four classes pleasant, but the rule has innumerable exceptions. Some people like garlic and some dislike nutmeg.

One reason that casual experience does not yield insight into the qualitative relationships of odors is that most olfactory stimuli arouse tactual or gustatory qualities, or both, in addition to their odors. Chloroform tastes sweet, and ether tastes bitter. Menthol is cool in the nostril, and vinegar produces a 'penetrating' tactual quality. A person anesthetic to smell can identify coffee by its 'prickle' and illuminating gas by its taste.

Stimulation. It has been estimated that there are about 60,000 different odorous substances, but the chief effect of many of these is tactual or gustatory. Perhaps there are not more than fifty from which the particles diffused in the air have no sensory effect other than the stimulation of the olfactory receptors. The total list of stimuli would contain many organic and inorganic compounds. Most chemical elements are odorless, but chlorine, bromine and iodine have smells. The large group of organic compounds which are derivatives from benzene contain so many odorous substances that the class has been called 'aromatic.'

Odorous substances liberate molecular particles into the air, either by evaporation or by chemical change, and the particles are spread about by gaseous diffusion. These particles are the true stimulus to smell. Odorous liquids in direct contact with the olfactory region, as when the nose is completely filled with eau de Cologne with the head inverted, do not arouse smell.

Investigators have long sought for a satisfactory chemistry of smell. It has been noted that odorous inorganic compounds generally involve elements from the fifth, sixth and seventh groups of the periodic system. A theory that goes further suggests that the different olfactory qualities depend upon the molecular patterns of substances that are derived from benzene

or remotely related to it. This theory is too complicated and admits of too many exceptions to be described here.

Fusion and compensation. The mixture of two odorous substances gives rise to a perceived *fusion*, which can be analyzed by attention into its two components—some mixtures quite easily, and others only with difficulty. The intimate fusions sometimes closely resemble the odors that come from single unmixed stimuli, but they are never quite the same. A mixture of pure red and pure yellow lights can give a color identical with the color from pure orange light, but the analogous case in smell is never fully realized. For instance, the odor of a mixture of eucalyptol and pyradine in a ratio of 9.5 to 1 resembles, but is not quite the same as, the odor of camphor.

Some mixtures show intensitive *compensation*. The resultant odor is much weaker than either of the odors of its separate components. Compensation is most effective when the antagonistic components are weak. It is never complete: it is not possible for two perceptible stimuli to mix so as to give an odorless resultant. The effective use of perfumes to combat the odors of the body, in the days when frequent bathing was less common than it is now, depended upon partial compensation. Both components persisted, but in a weak fusion.

Adaptation. Olfactory adaptation takes place fairly rapidly and may become complete in a few minutes. The time depends upon the nature and intensity of the stimulus. Measured times for complete adaptation to camphor have varied between 5 and 7 min.; to eau de Cologne, between 7 and 12 min.; to balsam, between 3 and 4 min.; and to tincture of iodine, about 4 min.

Fig. 47 shows how adaptation is dependent upon the stimulus substance and its intensity. Intensity is measured in olfacties (see p. 151 *infra*). The degree of adaptation is measured by the threshold, *i.e.*, the average minimal intensity of stimulus necessary to elicit the odor. The curves show that the threshold increases as adaptation continues, that adaptation is more rapid for benzoin than for rubber and that adaptation is more rapid

with either substance for the greater stimulus intensity (14 *vs.* 10 olfacties for rubber; 9 *vs.* 3.5 olfacties for benzoin).

Adaptation is *selective*. Sensitivity is reduced most for the stimulus to which adaptation has been taking place, but it may be somewhat reduced for other stimuli, and not at all for still others. Thus adaptation to camphor results in diminished sensitivity to cologne, clove and ether. Adaptation to ammonium sulphide reduces the effectiveness of hydrogen sulphide and bromine, but not of oils and coumarin. Oil of guaiacum seems to have a universal effect, reducing in some degree sensitivity to all substances. No one has yet succeeded in working out a significant classification of odors on the basis of these common relations under adaptation.

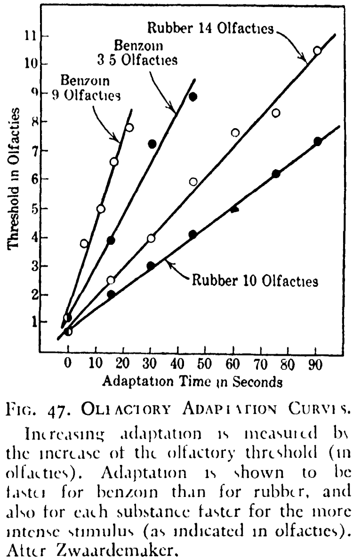


FIG. 47. OLFACTORY ADAPTATION CURVES. Increasing adaptation is measured by the increase of the olfactory threshold (in olfacties). Adaptation is shown to be faster for benzoin than for rubber, and also for each substance faster for the more intense stimulus (as indicated in olfacties). After Zwaardemaker.

Adaptation shows *modulation*, that is to say, the qualities of some odorous substances change as adaptation goes on. Thus at high intensities ionone resembles cedarwood, but at low intensities it is like violet; the quality of its odor changes as it gets fainter under adaptation. Modulation sometimes has a disastrous effect with cheap perfumes, which shift from a flowery odor toward the putrid as stimulation continues. Though pronounced with some stimuli, modulation is slight for others and non-existent for still others.

Recovery after complete adaptation seldom requires more than 5 min. The rate of recovery depends upon the time of adaptation and the intensity of the stimulus.

It is plain that adaptation plays an important role in everyday life. Man is very sensitive to odor, and some of the unpleasant odors, as of foul matter, are warnings useful in the preservation of health. However, after the warning is once given, adaptation takes place rapidly. Adaptation is a great advantage to garbage collectors and the workers in cheese factories. Unpleasant olfactory substances often have to be lived with. On the contrary, when the environment changes and the stimulus disappears, there is rapid recovery so that a new stimulus or the recurrence of the old is again as effective as before.

Sensitivity. Sensitivity is sometimes measured by the threshold: the greater the threshold the less the sensitivity. And for smell the threshold is measured in a variety of ways. We may determine for perception the least concentration of an olfactory substance in solution, or the least amount of diffused substance in a liter of air, or the least distance at which a given stimulus can be placed, or the least amount of surface of the olfactory substance over which the current of air into the nostrils is drawn. The *olfactometer* is based upon this last principle. The odorous substance is made the lining of a tube through which the air to the nostrils passes, and stimulus intensity is measured by the amount of exposed lining through which the air passes. The functioning length of the stimulus tube is varied by sliding the tube over and along another tube. This is the simple olfactometer for solids. A similar device is made for liquids, and there are other more elaborate forms of the instrument.

The unit of olfactory intensity is the *olfactie*, which is generally defined as the intensity of the smell of India rubber when the air is drawn through 0.7 cm. of India rubber tube in a standard olfactometer, *i.e.,* over a surface 1.76 sq. cm. This value was selected as the normal threshold for India rubber. When other odors are used, the olfactie may be determined experimentally as the area of the new stimulus required for the normal threshold.

To some substances the olfactory organ is very sensitive. The threshold value for mercaptan is about 0.000,000,000,000,043 gm. per cc. of air, and it has been estimated thus that one is able to perceive a mere 21,000,000,000 molecules of mercaptan. However, substances differ greatly from one another in effectiveness, and Table X shows certain typical thresholds. In this connection it should be remarked that the threshold is increased about four-fold when the observer is required to say exactly what odor it is that he perceives. Ordinarily he is required to report only whether he perceives some odor or no odor.

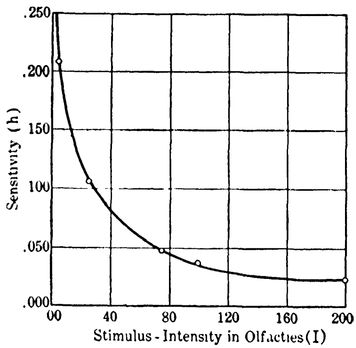


FIG. 48. SENSITIVITY CURVE.
Sensitivity is measured by h, the index of precision of the psychometric function; see pp. 54f. The curve is hyperbolic in form and is defined by the equation $hI^b = a$, where a and b are constants. The curve is for fresh rubber tubing known as 'laboratory red,' from unpublished data from an experiment by M J Zigler and A. H. Holway. See also (6)

TABLE X

OLFACTORY THRESHOLDS

The figures show for different substances the least concentration which on the average gives rise to a perceptible odor. From Zwaardemaker, after Passy.

Substance	Threshold Concentration Mm per liter of air
Methyl alcohol	0 30
Formic acid	0 40
Acetone	0 04
Camphor	0.000,0048
Ionone	0 000,000,095
Phenol	0 004
Pyradine	0.000,04
Methyl ether anthranilate	0 000,0006
Artificial musk	0 000,001

Fig. 48 shows the relation of sensitivity to the intensity of the olfactory stimulus. The curve is for red rubber. Sensitivity decreases as the absolute intensity increases, the familiar principle of the Weber function (see pp. 197-201).

Physiology. There is no satisfactory theory of the way in which olfactory receptors act selectively in mediating the different qualities of odor. There has been some speculation for which supporting facts are insufficient.

REFERENCES

Taste

1. Crozier, W. J. Chemoreception. In *Handbook of general experimental psychology*. Worcester, Mass.: Clark University Press, 1934. Pp. 987-1036.
2. Henning, H. Die Qualitatenreihe des Geschmacks. *Zsch. f. Psychol.*, 1916, 74, 203-219.
3. Von Skramlik, E. *Handbuch der Physiologie der niederen Sinne*, I: *Die Physiologie des Geruchs- und Geschmackssinnes*. Leipzig: 1926.

Smell

4. Henning, H. *Der Geruch*. Second edition. Leipzig: 1924.
5. Parker, G. H. *Smell, taste, and allied senses in the vertebrates*. Philadelphia: J. B. Lippincott Co., 1922.
6. Zigler, M. J., and Holway, A. H. Differential sensitivity as determined by amount of olfactory substance. *J. Gen. Psychol.*, 1935, 12, April.
7. Zwaardemaker, H. *L'odorat*. Paris: 1925.

See also Crozier (1), Von Skramlik (3)

CHAPTER 7

SOMESTHESIS

Every sensory experience that cannot be classed in one of the four specialized senses—sight, sound, smell and taste—is commonly ascribed to the sense of touch. This classification is a heritage from the past, impressed upon us by the doctrine of the five senses which has come down from antiquity. It has long been known, however, that 'touch,' the so-called fifth sense, is a congeries of unrelated sensory qualities and not a single group.

The sensory qualities popularly ascribed to touch derive, as we shall see, from a variety of organs and tissues: skin, muscles, tendons, ligaments, bones, joints and other internal organs. Since the structures they involve are so widely spread throughout the body (soma), they are fittingly called *somesthetic sensations.*

Our first task is to separate out the simple qualities from the miscellany normally given us by the body. This analysis may be achieved only by the isolation and separate stimulation of the different tissues. We must not accept as ultimate the blends, fusions and complexes which result from the simultaneous stimulation of several of these tissues. Our method, therefore, is clear. We must explore every one of the tissues in turn with all sorts and kinds of stimuli—mechanical, thermal, electrical, chemical—and describe and classify the experiences which result. We begin with the skin and shall take up the other tissues and organs in their turn.

CUTANEOUS QUALITIES

In searching for the qualities derived from the skin, the subcutaneous tissues must first be excluded from considera-

This chapter was written by K. M. Dallenbach of Cornell University.

tion. We must use stimuli that are too weak to excite them, although strong enough to excite the skin. With these stimuli we explore the surface of the skin, point by point; and we find that the skin is not uniformly sensitive, that only certain spots or points react and that the rest of the skin—by far the greater proportion—is insensitive. Thus the skin resembles a mosaic in which sensitive spots are inlaid haphazardly in an insensitive ground. The sensitive spots are specific in their responses and, if experimental conditions are rigidly controlled, fixed in their positions. They are of four kinds, and they yield four sensory qualities: pressure, pain, cold and warmth. We shall consider these qualities in order.

CUTANEOUS PRESSURE

Stimulus. Pressure is aroused by a deformation of the skin. The deformation may be either positive or negative; that is to say, the end-organs of pressure (Figs. 51, 52 and 53) respond to a pull as well as to a push. For example, if we pull a tuft of hair gently, or a thread glued to the skin, we find that the same quality is experienced as though we had depressed the skin. Any stimulus, in short, that gives rise to sufficient deformation of the skin will arouse a quality of pressure.

Quality. The qualities of pressure vary for different spots and for different stimulus intensities. If the hair of the skin is moved slightly or if the smoothly shaved skin is gently brushed with floss or a feather, a weak, ticklish, indefinitely localized quality appears. The same *tickle* may also be aroused at certain spots on the skin by weak punctiform stimuli. If such a punctiform stimulus is intense, spots are found which yield a solid, pressury quality, called *neutral pressure* to distinguish it from other pressures. If the stimulus is of intermediate intensity, there occur spots which give a weak, bright quality intermediate to tickle and neutral pressure. This quality is called *contact.* The correlation between intensity of stimulus and quality of experience is not exact. Different spots are attuned to different intensities; thus a stimulus that evokes

contact at one spot may arouse tickle or neutral pressure at another.

Pressure spots. The number and distribution of the pressure spots vary greatly for different areas of the body, and, for any given area, with the intensity and size of stimulus. When weak punctiform stimuli are used, the sensitive spots on the hairy regions of the body, comprising about 95 per cent. of the skin surface, are found to the 'windward' of the obliquely

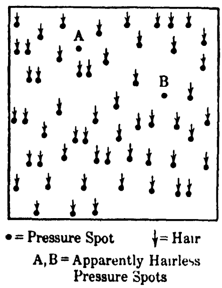


FIG. 49. MAP OF PRESSURE SPOTS AND HAIRS ON VOLAR SIDE OF MIDDLE FOREARM.
After Strughold.

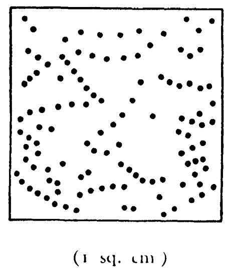


FIG. 50. MAP OF PRESSURE SPOTS FROM HAIRLESS REGION BETWEEN THUMB AND INDEX FINGER.
After Goldscheider.

emerging hairs (Figs. 49 and 51). On the hairless regions, like the palms of the hands and the soles of the feet, there are also definite points or spots of sensitivity (Fig. 50) which yield qualities similar to those derived from the hairy regions.

Some parts of the skin—*e.g.*, finger tips, lips and scalp—are highly sensitive and provide more than 100 pressure spots per sq. cm. Other areas—*e.g.*, the conjunctiva of the eye and parts of the mouth cavity—are insensitive to cutaneous pressure. Between these extremes may be ranged the remaining areas of the body as the figures for the following sample areas show.

Region	No.	Region	No.	Region	No.
Back	26	Elbow	12	Upper leg	14
Breast	22	Forearm	16	Lower leg	6
Upper arm	9	Back of hand	28	Top of foot	24
		Ball of thumb	135		

It will be noticed that the number of spots tends to increase toward the extremes of the limbs: there are more pressure spots on the hand and foot than upon the upper arm and leg. The average number of pressure spots per square centimeter of body surface has been estimated as about 25, but the particular deviations from it are large and numerous.

The number of pressure spots localized in any given area varies with intensity and size of stimulus. With an intense or large stimulus, every point or spot on the skin may yield pressure experiences, since the effect of the deformation caused by the stimulus may extend to a neighboring end-organ. If, as shown in Fig. 51, an intense stimulus is placed at *A* or *B*, positions not directly above a hair follicle, its effect may never-

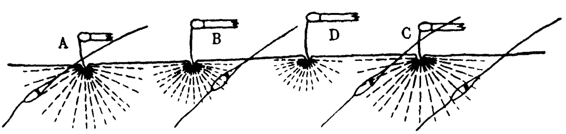


FIG. 51. THE EFFECT OF STIMULI OF DIFFERENT INTENSITIES WHEN APPLIED TO DIFFERENT SPOTS OF THE SKIN.

The stimuli are hairs which are pressed against the skin. The longer and finer the stimulus-hair, the less the pressure

theless spread to the end-organ and thus arouse an experience of pressure. If placed at *C*, between two neighboring hairs, it may similarly stimulate both of the end-organs. It is only when the stimulus is too weak for its effect to reach the end-organ, as at *D*, that no quality of pressure is experienced. If this weak stimulus is to be effective, it must be placed directly over the hair follicle, *i.e.*, to the windward of the hair.

Receptors. The close topographical relationship which obtains between the hairs of the skin and the pressure spots, together with the histological findings that the hair bulbs are supplied with highly specialized nerve endings (Fig. 52), leads inevitably to the conclusion that the hair follicles are the end-organs of pressure in the hairy regions of the body. The end-organs for the hairless regions, however, are not so defi-

nitely established. Meissner's corpuscles (Fig. 53) are generally assumed to function as such. They are the only specialized end-organs in the hairless regions occurring in sufficient numbers to be correlated with the pressure spots, and the numbers of spots and corpuscles vary together in different parts of the hand.

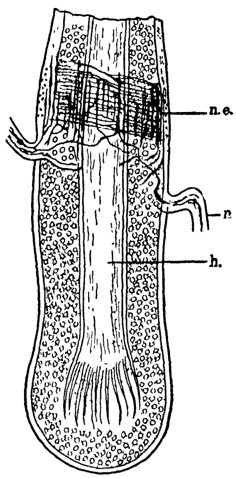


FIG. 52 THE SPECIFIC NERVE ENDINGS IN THE HAIR BULB.

n.e. = nerve endings, *n* = nerve, *h* = hair. After Bohm.

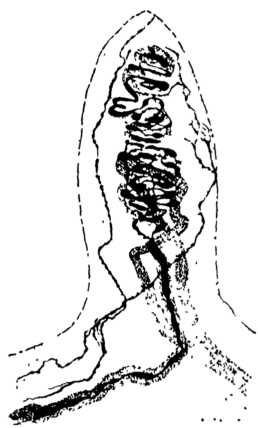


FIG 53. MEISSNER CORPUSCLE.
After Simonelli.

Sensitivity. The pressure spots differ from one area to another in sensitivity as well as in number. The threshold—expressed in terms of grams per millimeter, a unit obtained by dividing the pressure of the stimulus in grams by the average radius in millimeters—varies from 0.5 to 4 gm. per mm.

Region	gm./mm.	Region	gm./mm.	Region	gm./mm.
Back	4.0	Kneecap	1.93	Hand (dorsal)	1.27
Chest	2.7	Upper arm	1.44	Finger tips	0.57
Lower leg	2.16	Elbow	1.33	Lips	0.5

These figures indicate that sensitivity, like the frequency of the spots, tends to increase toward the extremities of the body.

The effect of the pressure stimulus upon any given area of the skin depends upon the weight of the stimulus object, its size and its speed of application. These factors vary concomitantly: weight and speed vary inversely, whereas weight and size vary directly.

The *differential limen, i.e.,* the average amount of change in the stimulus which is necessary to produce a just noticeable difference in sensation, varies greatly for different parts of the body. It is largest on the trunk, smaller on the head, considerably smaller on the face and smallest—falling very rapidly as advance is made peripherally from the trunk—at the extremes of the limbs.

Time relations. Sensations of pressure rise very rapidly after the application of the stimulus and fall very quickly after its removal. Their duration—probably the shortest of all our sensory experiences—varies for different parts of the skin surface from 0.01 to 0.0006 sec. These results have been obtained by noting how rapidly intermittent stimuli can be applied to the skin without causing the sensory experiences to fuse. If a rotating toothed wheel or, better, a hair attached to one of the prongs of an electrically driven tuning fork is brought into contact with a pressure spot, intermittence will be felt at frequencies varying, for different parts of the body, from 100 to 1552. Frequencies of about 100 fuse upon the upper arm and the tops of some of the toes. Frequencies of 1000 can be observed upon a few regions, *e.g.,* the finger tips, lips and tongue; and frequencies of 1552 only upon the tips of the index and great fingers.

The phenomenon of *adaptation* is most readily observed with weak stimuli. Then the pressure experience declines rapidly in intensity and disappears, although it sometimes recurs at the removal of the stimulus. Complete adaptation may also occur with intense stimuli. Instances from everyday life are numerous: we do not perceive the pressure of our clothes,

glasses, rings, etc. All that seems to be necessary for adaptation is prolonged and constant stimulation.

The after-effect of pressure is a positive after-sensation. This usually escapes notice, but such instances as failing to retrieve one's glasses or ring after washing probably point to the presence of the after-sensation.

Theory. The end-organs of pressure are the hair bulbs and the Meissner corpuscles. The questions with which we are now concerned are: (1) how are the end-organs thrown into function, and (2) what takes place within them when they function.

The end-organs are aroused, as we have seen, by a positive or negative deformation of the skin, but the presence of a liminal deformation is not alone sufficient to explain their stimulation. For example, if a hand is placed in a bucket of water, or a finger dipped into a bowl of mercury, pressure is not sensed over the immersed areas, where the positive deformation occurs, but only at the points of emergence. Again, if the skin is drawn into a glass tube or jar by means of a vacuum, pressure will be sensed, not over the entire area where the negative deformation occurs, but only along the edges. Since sensations occur only at points on the skin where differences in stimulus intensities obtain, the conclusion seems inescapable that the receptors are excited, not by the physical pressure itself, but by differences or gradients in pressure.

This theory, known as the theory of the pressure gradient, accords with all the known facts. The skin is composed principally of water, which is incompressible; consequently, if a force is applied to the skin the physical pressure will extend in all directions. This pressure will be greatest at the point of contact, and will diminish rapidly as the distance increases vertically into the skin or laterally along its surface. A steep gradient is correlated with a strong stimulus, a gentle gradient with a weak stimulus. The pitch of the gradient and the rate of its formation vary with the weight, size and speed of the stimulus.

It is not possible, at the present state of our knowledge, to say just what happens in the end-organ when the pressure gradient is formed. Two plausible suggestions, however, have been made. Both rest upon the fact that the energy of the adequate stimulus for the end-organ is very small (about 0.002 erg), being only about a thousandth of the energy required to stimulate the nerve directly. The first suggestion is that the end-organ acts like the receiving set of the radio, picking up the weak energy released by stimulation, and transforming it into a neural impulse. The second suggestion is that the end-organ acts as a condenser: the external stimulus, though weak, is sufficiently strong to discharge the energy stored in the end-organ and thus indirectly to initiate a neural impulse.

CUTANEOUS PAIN

Pain, once regarded as a common sensation produced by excessive stimulation of any sensory nerve or end-organ, is now known to be a separate sense with its own qualities and specific sense organs. With few exceptions, pain is aroused in all tissues and organs of the body. For the present we shall limit the discussion to cutaneous pain.

Stimulus. Cutaneous pain is aroused by various classes of stimuli: mechanical (sharp as well as dull points), thermal (extremes of physical heat), electrical and chemical. Any stimulus, in short, that is intense enough to injure the free nerve endings in the skin will arouse pain.

The most suitable pain stimuli are sharply pointed hairs, bristles, thistle spines or fine needles. They are held in an instrument in which the weight that presses the point against the skin can be controlled.

Unless the skin is especially prepared, pain is not the first quality aroused when the stimulus point is applied. Pressure normally precedes pain and blends with it. This result is due to the peculiar structure of the skin. The outer layer, the

epidermis, is hard and horny; the underlying layer, the cutis, is soft and spongy. Under ordinary circumstances, the impact or jar caused by the application of the stimulator is transmitted to the pressure end-organs imbedded in the cutis before the epidermis has been pierced. The more superficial end-organs of pain are not affected by this impact. If the horny layer of the skin is removed, either by being softened with soap and water and scraped away, or by being blistered and cut away, and if the tissues thus prepared are kept soft and moist by sponges wet with a salt solution, pain may be aroused earlier than pressure and quite independently of it.

Quality. Language is particularly rich in terms descriptive of the experiences of pain: achy, beating, biting, boring, bright, burning, clear, cutting, dark, digging, dragging, drawing, dull, fluttering, gnawing, hard, heavy, itchy, nipping, palpitating, penetrating, piercing, pinching, pressing, pricking, quick, quivering, radiating, raking, savage, sharp, smarting, squeezing, stabbing, sticking, stinging, tearing, thrilling, throbbing, thrusting, tugging, twitching, ugly, vicious. Most of these terms, however, characterize (1) the temporal course of the experience—*e.g.,* palpitating, throbbing; (2) its spatial distribution—*e.g.,* penetrating, radiating; (3) its fusion or integration with pressure—*e.g.,* heavy, pressing; or (4) its effective coloring—*e.g.,* savage, ugly. Only a few of the terms are purely qualitative, *viz.,* achy, bright, clear, dull, itchy, pricking, and quick; and only a few of these characterize the *qualities* derived from the skin itself.

The qualities of pain extend from *itch,* through *prick,* to a clear, painful quality, called *clear pain.* The qualities aroused vary with the spots stimulated and the intensity of the stimulus. In general, weak intensities elicit bright, itchy qualities and more intense stimuli wiry pricks and clear pains.

Any of these qualities of pain may be either pleasant or unpleasant according to circumstances. The itch of a healing wound, the prick of spice upon the tongue or the clear pain of a loose tooth which a child wiggles and pries with its tongue

is pleasant. In other situations the same qualities may be indifferent or unpleasant.

In addition to these qualities, which lie within the itch, prick, clear-pain continuum, certain areas of the skin—*e.g.*, under the finger nails and in the auditory meatus—yield a bright, lively quality called *quick pain*. This quality is sometimes described as ugly or vicious, and, unlike the other qualities derived from the skin, is inherently unpleasant.

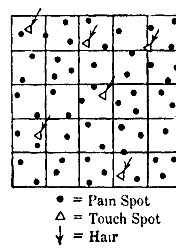


FIG. 54. MAP OF PAIN AND TOUCH SPOTS AND OF HAIRS. 1/4 sq cm on volar side of forearm After Strughold

Pain spots. The pain spots are much more numerous than the touch spots. In Fig. 54 the ratio is about 7 : 1. The *distribution* of the pain spots is haphazard and unrelated to that of the touch spots or the hairs. The frequency of distribution varies considerably in different parts of the body. A few regions are insensitive to pain, like the insides of the cheeks opposite the second lower molars; other areas vary widely in the number of pain spots found within them, as the figures (no. per sq. cm.) for the following sample areas show.

Region	No	Region	No	Region	No
Tip of nose	44	Scalp	144	Inguinal region	220
Sole of foot	48	Eyelid	172	Elbow inner joint	224
Ball of thumb	60	Back of hand	188	Neck jugular region	228
Edge of ear	76	Back of forearm	196	Knee inner joint	232

These results show that the pain spots are most numerous where the large nerves and blood vessels come nearest to the surface of the skin. Their average number per square centimeter of body surface is estimated at about 175. This value, however, is of little significance, since the deviations, as indicated in the table just above, are both large and many.

Receptors. The free nerve endings in the epidermis, shown in Fig. 55, are identified by most investigators as the receptors that function in the arousal of cutaneous pain. The evidence, though indirect, is fairly conclusive. The free end-

ings are the only nerve structure in the epidermis; and they alone are found in sufficient numbers to tally with the number of pain spots. Moreover, they are the only structures found in areas which are exclusively sensitive to pain—*e.g.*, the cornea—and they are absent from areas that are completely insensitive to pain—*e.g.*, the mucous membrane of the cheeks. Further evidence for the correlation between pain and the free nerve endings is obtained from the use of local anesthetics. If, for example, an oleic acid solution of cocain, which acts only on the superficial nerve endings, is rubbed into the skin, only the sense of pain is lost, while sensitivity to pressure, to warmth, and to cold is unimpaired.

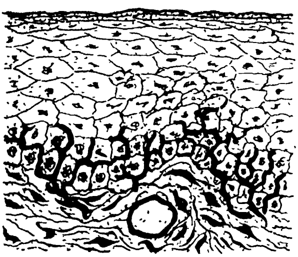


FIG. 55. FREE NERVE ENDINGS IN THE EPIDERMIS. After Kadanoff.

Sensitivity. The thresholds of the pain spots vary considerably in different parts of the body, but the deviation among adjoining spots is slight. The following table shows representative values, expressed in terms of the weight upon the stimulus-point.

Region	gm.	Region	gm.	Region	gm.
Tip of nose	1.00	Scalp	1.35	Inguinal region	0.38
Sole of foot	2.40	Eyelid	0.18	Elbow inner joint	0.27
Ball of thumb	2.30	Back of hand	0.66	Neck jugular region	0.38
Edge of ear	0.78	Back of forearm	0.51	Knee inner joint	0.30

These results show that regions with the most pain spots have as a rule the smallest thresholds, although there are exceptions. Possibly the thickness of the epidermis affects the threshold. Sensitivity to pain, whether measured by number of spots or by the threshold, does not increase, as pressure sensitivity does, toward the extremes of the limbs. Sensitivity to pain is greatest in the eyelids, ears and places where the large nerves and blood vessels come near the surface of the skin in 'vital' areas

of the body, the protection of which is of biological significance.

Time relations. The *rise and fall of pain*, from the application until after the removal of the stimulus, are slow and gradual. Although the times vary considerably, they tend, nevertheless, to be longer for pain than for any other sense. Instances illustrating the slow rise of pain are common in everyday life. If one stubs a toe, pressure is sensed appreciably earlier than pain; if one touches a burning hot object, pressure and temperature experiences come first. Experiments merely make these observations more precise. Cutaneous pain rises in from 0.03 to 1.00 sec. The disappearance of pain is usually slower than its rise.

Recent experiments show that pain undergoes *adaptation*. Complete adaptation can be reached if the stimulus is of unvarying intensity. The painful stimuli of everyday life are seldom constant. The time of adaptation varies considerably (from 6 sec., to about 12 min.) with the observer, with the nature of the stimulus (mechanical, thermal, etc.) and with the spot stimulated. It does not appear to depend upon the intensity of stimulus.

When a painful stimulus is removed, there follows an after-pain whose intensity and duration are correlated with intensity of stimulus.

Theory. The free nerve endings of the epidermis are, as we have seen, the pain receptors. Any injury or rupture of their membranes is accompanied by changes in the tissue-solutions. The osmotic pressure gradient thus formed is believed to be the immediate stimulus which excites the nerve fiber and initiates the sensory impulse. This gradient is sluggish; it changes slowly. The long latent period and after-pain imply such a gradient. This theory also explains why pain may be sensed at one point and not at an immediately adjacent point. The osmotic pressure gradient is effective only when a free nerve ending is ruptured. No matter how close a stimulus

may be to a nerve ending, it will not arouse pain unless the nerve fiber itself has been injured.

CUTANEOUS COLD AND WARMTH

Though cold and warmth are qualitatively as dissimilar as any two senses, we nevertheless consider them together here because both are aroused by similar stimuli.

Stimulus. The direct stimulus for cold and warmth is a change in the temperature of the receptors and not the outside molecular and radiant energies known physically as heat. Thus, a given temperature may, according to circumstances, be cold, indifferent, warm or hot. Moreover, only a small range of the intensities of physical heat—that extending from about $-10°$ to $+70°$ C.—gives rise to temperature experiences.

Temperatures below and above this narrow range usually evoke only pain (p. 168), and those falling within it tend to evoke cold or warmth or nothing at all, accordingly as they affect the temperature of the skin and the heat flow which normally passes outward from it. As long as the outward flow of heat, which is maintained by the blood stream, is normal and constant, cold or warmth is not aroused—we are indifferent as regards temperature. If, however, the heat flow is increased and the temperature of the skin falls, cold is evoked; if the outward flow is decreased below the normal, or if it is reversed so that the flow is inward and the temperature of the receptor rises, warmth is evoked.

The temperature of the skin at the point of indifference is known as *physiological zero.* The value of this constant varies in different parts of the body. It also varies in the same part of the body at different times and with such factors as one's health, activities, surroundings and mental state. When a normal and healthy person is resting or moving quietly in a room at 20° to 22° C., physiological zero is about 37° C. under his tongue, 35° on the normally clothed parts of his body, 33° on his hands and face and about 26° on the ear-lobes in which the blood stream flows sluggishly. That the

physiological zero varies markedly in different parts of the body is quickly demonstrated by grasping an earlobe between thumb and fingers. The earlobe is cold to the fingers, and the fingers are warm to the earlobe, although each region by itself is thermally neutral.

Objects with temperatures above physiological zero usually arouse warmth or heat, and those with temperatures below that value, coolness or cold. The quality aroused, however, does not depend entirely upon (1) the object's temperature in relation to physiological zero, but also upon (2) its specific heat, (3) its thermal conductivity, (4) its size, (5) the nature of its contact with the skin and (6) the state of adaptation of the sense organ.

If the skin is touched with objects differing in specific heat and thermal conductivity (*e.g.*, wool, cotton, cork, wood, oil-silk, metal, glass) but having the same temperature—let us say that of the room (22° C.)—they are not neutral, like the air of the room. Some of them are warm (wool, cotton), others cold (metal, glass), while still others are first cool, then indifferent and then mildly warm (wood, oil-silk, cork). Wool is warm because, while it conducts heat as poorly as air, it checks the flow of heat from the part of the skin covered, so that the temperature of the skin and of the underlying receptors rises. Metal, on the other hand, is cold because it conducts heat much better than air, and it draws more heat from the skin, so that the temperature of the skin and of its receptors falls.

Cold spots, normally stimulated by temperatures below physiological zero, are also stimulated by certain temperatures above it. For example, a cold spot, in a region of the skin whose physiological zero is 33° C., gives no response to temperatures ranging from 34° to 43°, but to temperatures above about 43° it responds with a distinct and intense experience of cold. This response is called *paradoxical cold*, since it seems paradoxical that a warm object should arouse cold. Warm spots respond with *paradoxical warmth* when stimulated with temperatures below physiological zero between about 25° and

31°. These responses, however, are weaker and much less insistent than the paradoxical responses of the cold spots.

Quality. There are few if any qualitative differences of cold and warmth: cool and tepid are probably merely intensitive gradations; biting or stinging cold, and burning or pricking heat, are integrations of pain and cold, and pain and heat, respectively.

Besides cold and warmth, a third quality, *heat*, has to be added to our list. It resists analysis, and it resembles neither cold nor warmth from the excitation of whose receptors it is derived; it must, therefore, be regarded as a unique quality. It is aroused by the simultaneous stimulation of the warm and

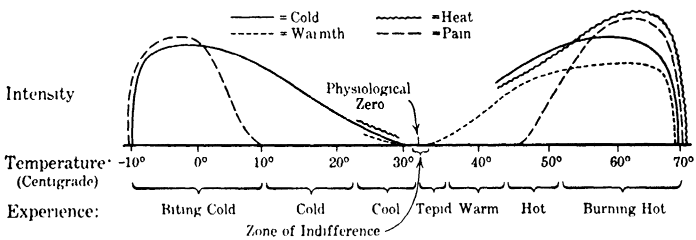


FIG. 56. CORRELATION BETWEEN STIMULUS-TEMPERATURE AND SENSATION.

cold spots, whether the stimulation is normal or paradoxical. For example, heat is aroused by areal or simultaneous punctiform stimulation of a cold and warm spot (1) with temperatures of 45° C. and above (normal warmth and paradoxical cold); (2) with two temperatures each of which alone is inadequate to heat (*e.g.*, 20° C. on a cold spot and 38° C. on a warm spot); and under favorable circumstances (3) with temperatures of 25° to 28° C. (paradoxical warmth and normal cold).

The relation between stimulus temperature and the qualities aroused for an area of the skin with a physiological zero of 33° C. is shown in Fig. 56.

Cold and warm spots. The cold and warm spots are found, like the pressure and pain spots (pp. 156, 163), over practically the entire surface of the body. They differ greatly, how-

ever, from each other in number and distribution, both on the whole and in the different parts of the body. The number of cold spots varies from about 19 per sq. cm. in the upper lip to about 1 to 2 on the palmar side of the fingers. Between these extremes the distribution is as follows:

Region	No	Region	No	Region	No	Region	No.
Lower lip	16.0	Abdomen	12.5	Upper arm	5.0	Thigh (front)	5.2
Nose (tip)	13.0	Lower chest	10.2	Elbow	6.5	Calf	4.3
Chin	9.0	Upper chest	9.0	Lower arm	7.5	Shin	5.7
Cheek	8.5	Loin	8.0	Hand (back)	7.4	Foot (top)	5.6
Forehead	8.0	Back	7.8	Hand (palm)	1.0–3.0	Foot (sole)	3.4

The average frequency of cold spots is estimated at about 7 per sq. cm. of body surface. This average is significant, since the values for the different regions do not deviate greatly from it.

The cold spots have a characteristic distribution shown in Fig. 58, and are more or less uniformly scattered over the

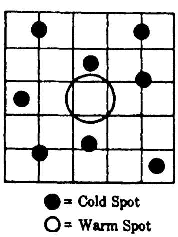


Fig. 57. Composite Map, by Continuous Exploration, of an Area on the Back of the Hand

Area = 1 sq. cm

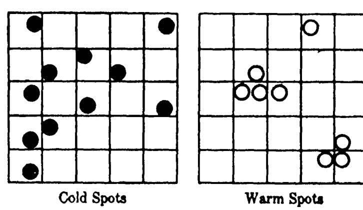


Fig. 58. Separate Maps, by Punctiform Exploration, of an Area on the Upper Arm.

Area = 1 sq. cm.

skin. The warm spots, on the contrary, vary considerably in number and distribution with the method of localization. When the exploring stimulus is moved continuously the spots are found to be large and diffuse, few in number, and widely separated (see Fig. 57). Their number varies from about 2 per sq. cm. on the sides of the fingers to 0.2 on the inner surface of the upper arm. The numbers in other areas are as follows:

Region	Av.	Region	Av.	Region	Av.	Region	Av.
Cheeks	1.70	Breast	0.30	Elbow	0.70	Hand (palm)	0.45
Nose	1.00	Upper leg	0.39	Forearm	0.40	Finger (back)	1.75
Forehead	0.62	Upper arm (outer)	0.30	Hand (back)	0.54	Finger (front)	1.60

When exploration is punctiform the spots appear as small and definite, equaling or exceeding the cold spots in number, and grouped into clusters (see Fig. 58).

Receptors. The end-organs of cold and warmth are still unknown despite repeated efforts to identify them. The more direct method is histological. The tissue under a cold or warm spot is excised, cut into thin sections, stained and examined microscopically. This method has revealed no specialized end-organs but rather numerous nerves and undifferentiated nerve endings. It has therefore been supposed that the undifferentiated nerve endings are the receptors whose stimulation gives rise to cold and warmth.

Sensitivity. Very small changes in the temperature of the skin evoke noticeable differences in cold and warmth. The amount or degree of change required to arouse a just noticeable experience (threshold) or a just noticeable difference (differential threshold) depends upon (1) the place and size of the area stimulated, (2) its physiological zero or adaptation temperature and (3) the method employed, *i.e.,* whether the stimuli are applied successively to the same part of the skin or simultaneously to different parts.

The *absolute thresholds* of cold and warmth vary considerably in different parts of the body. The two thresholds are approximately equal for a given region of the skin, and are exceedingly small. They vary, with different adaptation temperatures of the fingers, from 0.033° to 0.088° for cold and from 0.038° to 0.081° for warmth. These thresholds are smallest, on the fingers, at temperatures of about 28°—approximately 4° below normal physiological zero—and increase slowly on each side of this point, as the data given below show.

Temperature	Cold	Warm	Temperature	Cold	Warm
16°	−0.076	+0.067	32°	−0.038	+0.041
20°	−0.059	+0.053	36°	−0.055	+0.062
24°	−0.042	+0.039	40°	−0.075	+0.081
28°	−0.033	+0.038	44°	−0.088	+0.078

The *differential limens* greatly exceed the absolute thresholds in size. Like the latter, however, they vary with the region of the skin and are approximately equal for cold and warmth at any given place. Local variation in differential limen is shown by the following results:

Region	D.L.	Region	D.L.	Region	D.L.
Fingers	0.2°	Temple	0.3–0.4°	Nipple	0.8°
Cheek	0.2–0.3°	Chest (sides)	0.4°	Back (sides)	0.9°
Hand (back)	0.3°	Abdomen	0.5°	Back (middle)	1.2°

Time relations. Cold, like pressure, rises rapidly after the application of the stimulus, and comes quickly to its full intensity; warmth, like pain, rises slowly and comes gradually to full intensity. These latent periods, however, vary in different parts of the body; they are short where the skin is thin, and long where the skin is thick. The difference in the latent periods of cold and warmth is responsible for the momentary flash of cold when one passes one's hand rapidly through hot water or steps under a hot shower. The cold receptors, because of their shorter latent period, are aroused (paradoxically) first, and cold does not give way to heat until the delayed excitation of the warm receptors is added.

Complete *adaptation*, or thermal indifference, can be realized with punctiform stimuli and with areal stimuli having temperatures close to physiological zero. With punctiform stimuli, the adaptation time varies considerably for different spots (from 4 to 97 sec. for cold, and from 2 to 68 sec. for warm), and it seems to depend upon the sensitivity of the spot rather than upon the temperature of the stimulus.

With other areal stimuli complete adaptation is not attained. One may, for example, have cold hands or feet for hours at a time. A certain degree of adaptation in the sense of blunting or dulling occurs, however, rather rapidly. The unpleasantly

cold temperature, for example, of the first plunge in a pool is quickly blunted if the swimmer persists; it may never reach thermal indifference, but it very soon becomes less intense and it may, as experience bears frequent witness, become pleasantly cool.

Theory. There is for every area of the skin an outward flow of heat, a negative thermal gradient, which is normal and constant for it. The heat flow on the dorsal forearm, for example, is about —0.01 cal. per sec. per sq. cm. When this gradient obtains, the forearm is indifferent as regards temperature, and neither warm nor cold spots are aroused. If the gradient, however, is made more negative, cold will be evoked at the cold spots, and within very definite limits warmth will be evoked paradoxically at the warm spots; the intensity of the experience will depend upon the steepness of the gradient and the speed of its formation. If, on the other hand, the gradient is decreased, or its sign reversed, marking an inward flow of heat, warmth is evoked at the warm spots. The cold spots remain inactive until a gradient of +0.1 cal. or greater is formed; then they respond paradoxically, in the only way that they can respond, with cold. The intensity of the experience, as before, depends upon the steepness of the gradient formed.

The steepness of the thermal gradient varies directly with the size of the difference between the temperature of the stimulus and the point of the physiological zero. When the differential between these temperatures is large, the gradient is steep and the resulting sensation is intense; when the differential is small, the gradient is slight and sensation is weak.

SENSIBILITY OF THE MUCOUS MEMBRANE

When the alimentary tract and its orifices are explored by methods similar to those employed for the skin, no new qualities are found excepting, of course, taste and smell. The marginal zones of the orifices are highly sensitive and yield, at definite spots or points, all the qualities with which we are familiar: pressure, pain, cold, warmth and (under simul-

taneous stimulation of cold and warmth) heat. Some cavities are insensitive. Others, such as the buccal and nasal cavities, respond to stimulation, much like the skin.

Punctiform stimulation is not feasible for exploring the alimentary tract. Special methods and devices must be used: stomach and rectal tubes having electrical heating coils, elastic balloons or electrodes attached to their inserted ends.

Except for the intestines, which cannot be reached experimentally, the entire alimentary tract yields pressure when its walls are distended by the inflated balloon; increased distention gives pain. Pain is also evoked by extremes of temperature and in some places by certain chemical stimuli. Both pressure and pain are aroused electrically by faradic stimulation. These qualities are deader and duller than their cutaneous counterparts, resembling more closely the pressures and pains derived from the subcutaneous tissues (see below). In all probability, the receptors are not situated in the mucous membranes of the tract but in its muscular walls.

The esophagus and the stomach yield also diffuse and faint experiences of cold, warmth and heat, which presumably derive from the mucous membrane. Hunger, thirst and nausea cannot be obtained through experimental stimulation of the alimentary tract. They are not simple qualities, but, as we shall note later (p. 186), complex perceptual experiences.

Nothing is known of the end-organs in the alimentary tract.

KINESTHESIS

The sensitive motor apparatus of the body—the muscles, tendons, ligaments, bones, cartilages and membranes of the joints—furnishes a mass of sensation which is difficult to analyze correctly into proper components. This system of sensory qualities is often called *kinesthesis*, because it forms the basis for the perception of bodily movement. It is also referred to as *proprioception*, because it provides the necessary afferent impulses which enable the organism to adjust its movements accurately (see pp. 13, 239f.). It is now known that kinesthesis derives from stimulation in *muscles, tendons* and *joints*, and we shall take up these three sources successively.

Muscles. The muscles can be isolated for experimental investigation by anesthetizing the skin above them and by passing an electric current through the joints and tendons to render them partially insensitive. Then we find that mild stimuli (weights, weak electrical currents) arouse muscular pressures that are characterized as dull, dead and diffuse, like the pressures in the alimentary tract. Strong stimuli (heavy weights, strong electrical currents, penetration by sharp points) arouse dull pain and, at great intensities, ache. These muscular qualities are called respectively *dull pressure, dull pain* and *ache.* Once we know what to look for, they can easily be observed when the muscles are contracted by active or passive movements. Another quality of pressure, a light, flashing, thrilling or *bright pressure*, appears at times under active contraction of the skeletal muscles. We can notice it easily in the calves of the legs and the thighs when we feel particularly well and are full of the vim and vigor of life, or when we are pleasantly excited.

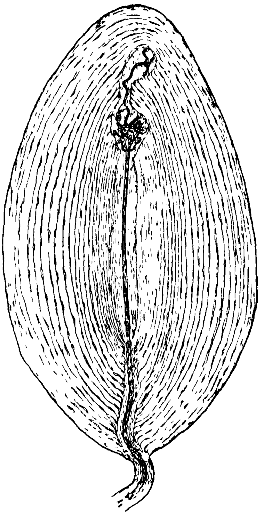


FIG. 59. PACINI CORPUSCLE. After Ranvier.

The muscles are abundantly supplied with sensory nerves and receptors. Free nerve endings terminate between the muscle fibers; Pacini corpuscles (Fig. 59) are found in the muscle sheaths, and muscle spindles (Fig. 60) lie within the connective tissue surrounding bundles of muscle fibers. It is probable, from what is known of the distribution of these structures, that the muscle spindles mediate dull pressure and that the free nerve endings, at different intensities of stimulation, mediate dull pain and ache. Pacini corpuscles present an enigma. They are found in tissues widely distributed throughout the body: in the skin, muscles, tendons, periosteum, joint capsules,

pancreas, peritoneum, mesentery, pleura and oviducts. It is possible that they are pressure organs, mediating bright pressure in the muscles and some other kind of pressure elsewhere.

Tendons. The tendons cannot be stimulated independently. The effect of the excitation is always likely to spread. The best we can do is to see what there is in kinesthesis that is not already attributed to the skin and the muscles. Thus we find *strain*, which, as the intensity of stimulation is increased, gives way first to *dull pain* and then to *ache*. These qualities may be observed, once they are identified, whenever we actively and vigorously exert ourselves (wrestling, lifting a heavy weight) or even better when we passively sustain a heavy weight.

Except at their junctions with the muscle, the tendons are sparsely supplied with receptors. These include free nerve endings, Pacini corpuscles (Fig. 59), and tendon spindles (Fig. 61). Strain is probably mediated by the tendon spindles, and dull pain and ache by the free nerve endings.

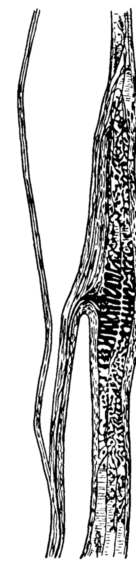


FIG. 60. MUSCLE SPINDLE. After Ruffini.

Joints. The tissues at the joints are so numerous and so closely packed together that we cannot hope to isolate and stimulate them separately. Movement of the elbow, knee and other joints ordinarily does not present a new quality. When, however, we use a joint that is rarely moved independently (*e.g.*, the first joint of the index finger), or when we move one in an unusual direction or manner (*e.g.*, the sidewise movement of the wrist with fingers extended), or when we increase the pressure upon the articular surfaces (*e.g.*, by moving a finger while it is strongly pressed into its socket), a soft, smooth pressure quality is evoked. This is called *smooth pressure* to distinguish it from the other qualities of pressure.

Smooth pressure is particularly clear when the fingers are

placed in the position shown in Fig. 62. The forefinger is clinched, the other fingers extended, and the first joint of the forefinger, which is pendant, is moved slowly back and forth.

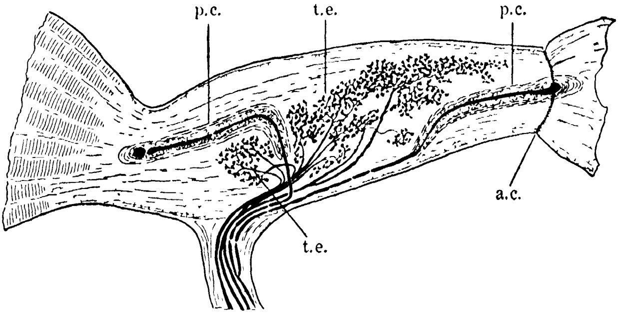


FIG. 61. TENDON-SPINDLE CONTAINING TWO MODIFIED PACINI CORPUSCLES. *p.c.*, Pacini corpuscle; *t.e.*, terminal expansion of end-organ; *a.c.*, annular constriction After Ruffini

The pressure is clearest when the observer has his eyes closed, and the movement is passive, *i.e.,* the finger is manipulated by another person. Qualities from muscles are then entirely lacking, and those from the skin are minimal.

FIG. 62. PROCEDURE FOR ELICITING THE QUALITY OF SMOOTH PRESSURE.

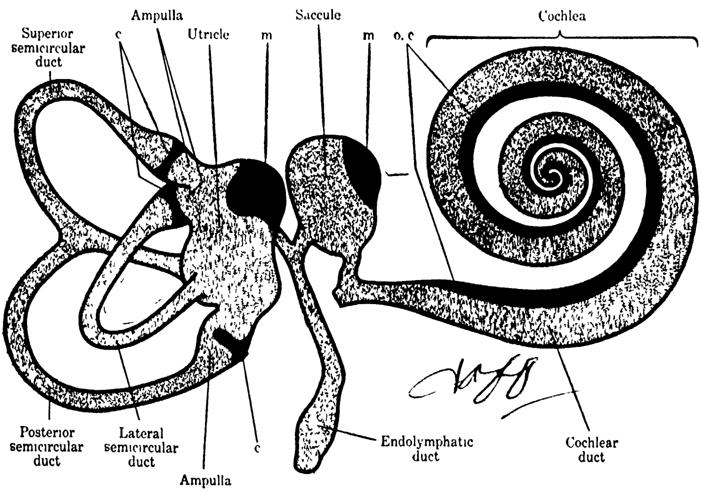
With the exception of the cartilages, which lack nerves, free nerve endings are found in all the articular tissues. They probably act here, as elsewhere, as pain receptors. Pacini corpuscles are less widely distributed, appearing chiefly in the synovial membrane. They probably mediate the smooth pressures.

SOMESTHESIS FROM THE INNER EAR

The labyrinth of each internal ear (Fig. 63) is made up of three parts: (1) the cochlea with its organs of Corti (see pp. 109f.), (2) the semicircular canals with their cristae, and (3) the saccule and utricle of the vestibule with their maculae.

We are now concerned only with (2) and (3). These are the most inaccessible tissues and organs that we have thus far considered.

Stimulation. The normal stimulus for the non-acoustical organs of the labyrinth is a change in the position of the head or a change in the acceleration or direction of the movement



Black areas are those with nerve structures
m = macula c = crista
o c = organs of Corti

FIG. 63. MEMBRANOUS LABYRINTH.

of the head, either circular or rectilinear. Circular movement probably affects the *cristae* of the *semicircular canals*; rectilinear movement and change in position of the head act upon the *vestibular* organs—the *maculae* of the *utricle* and *saccule.*

The receptors in *the semicircular canals* may also be stimulated directly in a number of other ways: (*a*) electrically, by passing current through the mastoid regions of the head; (*b*) thermally, by irrigating the external ear with hot or cold water. When the eardrum is missing, it is possible to induce stimulation (*c*) mechanically by pressing or pricking the canals with dull or sharp points, and (*d*) chemically by apply-

ing certain drugs to them. They may also be stimulated (*e*) indirectly or reflexly by excitations from the eyes and stomach, and by the indirect action of drugs. This indirect effect may result when one is looking down from great heights, or looking at the rotating walls of a room (the 'haunted swing' of many amusement parks), or looking at stationary walls while one is himself rotating. Again, an upset stomach—as in extreme nausea caused by the first cigar, by ipecac or by apomorphine—or the action of certain drugs, like alcohol, may have the same effect. That these indirect means are actually effective is supported by experiments upon deaf persons, most of whom lack functional semicircular canals. Great heights, the 'haunted swing,' etc., do not affect the deaf as they do the normal person. Such a difference in reaction can best be explained by difference in structure.

The following effects are obtained from the stimulation of the semicircular canals: (1) involuntary muscular reactions, such as nystagmus (a rapid, involuntary oscillation of the eyes), and adjustments of the head and limbs toward the maintenance of equilibrium; (2) widespread visceral and autonomic changes, *e.g.,* tightening of the abdominal walls, reversed peristalsis in the esophagus, rapid flow of saliva, excessive perspiring and blushing; (3) a varied mass of kinesthetic, cutaneous and other experiences, *e.g.,* pressures around the neck, face and eyes, cold chills, diffuse warmths, bad tastes, hints of disagreeable odors and visceral tensions, and (4) numerous perceptual experiences—a swimming sensation, perceptions of bodily rotation, dizziness or vertigo, nausea and visual illusions in which stationary objects appear to move and float in space.

The known effects of *vestibular stimulation* are meager. Most authorities claim that they are entirely unconscious reflexes, although others have held that, in addition to the reflexes, a pressury quality is aroused. These reflexes provide sources of stimulation that contribute to our perception of the position of the head and probably of the rectilinear movements

of the head and body. The resulting perceptions are discussed presently (see pp. 239-241).

We are under no constraint to find qualitative correlates for the non-acoustical organs of the inner ear. Other highly organized motor activities, *e.g.*, respiration and the pulsations of the heart, are maintained without the arousal of specific conscious qualities. It is not inconceivable, therefore, that bodily equilibrium may also be maintained unconsciously. The point of view that these organs have no sensory function but are merely afferent reflex mechanisms is strengthened by the fact that the vestibular nerve, which carries their impulses, has no path to the cerebral cortex. Some of its central branches run to the vestibular nuclei in the medulla, from which there is no known tract to the thalamus; others go without interruption to the cerebellum.

Time relations. The effects of stimulation of the semicircular canals appear with positive or negative changes in the *acceleration* of rotation of the head and disappear when the speed becomes *constant*. Their rise and fall, in comparison with the other sensory effects which we have considered, is slow; the times involve whole seconds. Furthermore, the different effects, *i.e.*, muscular, autonomic, sensory and perceptual, rise and fall at different times, and the times of all vary with the intensity and duration of stimulation.

Prolonged variable stimulation of the semicircular canals leads slowly, or not at all, to complete adaptation. What is a mere matter of seconds in the skin, is at least a matter of minutes or hours in the canals. Some people become adapted to the motion and churning of a boat, for example, in a short time; others may 'feel' the effects for as long a time as they are on the boat, and still others for hours afterward.

Periodic stimulation, on the other hand, may radically alter the experiences aroused by the semicircular canals. The intensity and complexity of the responses associated with the canals may be reduced and the various components may gradually drop out. Thus people whose occupations or professions

occasion repeated stimulation of the canals—*e.g.*, sailors, dancers, whirling dervishes, acrobats and stunt flyers—do not experience the effects which are normal to other persons.

The *after-effects* of stimulation of the canals are as complex and varied as the primary effects. If the canals are stimulated by way of rotation, for example, almost all the phenomena aroused continue unabated after the speed of rotation is decreased or the rotation is stopped. The most striking of these phenomena is the illusion of *reversed rotation*. When the objective movement is slowed or stopped, the observer perceives himself to be spinning around in the opposite direction. If the rotation or the swimming sensation circles clockwise during the objective movement, it will appear to circle counterclockwise when the speed of rotation is diminished. This phenomenon has been called the negative after-sensation of movement, but it is clear that it is a perception that depends directly upon stimulation in the normal manner. It results from a change in the acceleration of the objective movement, the kind of change that always affects the receptors in the canals.

Little is known about the temporal relations of the effects aroused by vestibular stimulation. The effects appear to rise and fall rapidly with changes in the position of the head and with positive or negative changes in the acceleration of rectilinear movement, and to adapt themselves quickly after the change is made or when the acceleration becomes uniform. There is no evidence, however, that they show adaptation to prolonged stimulation, or that they become dulled or obtuse with periodic stimulation. Elevator men, for example, do not appear to suffer a diminution in the sensitivity of their vestibular organs, or to show any other effects which can be attributed to the repeated, periodic stimulation occasioned by their occupation.

As in the stimulation of the semicircular canals, when the objective movement is slowed or suddenly stopped an illusion arises of movement in the opposite direction. This is most

noticeable in the vertical plane, but even here it is weak and of short duration.

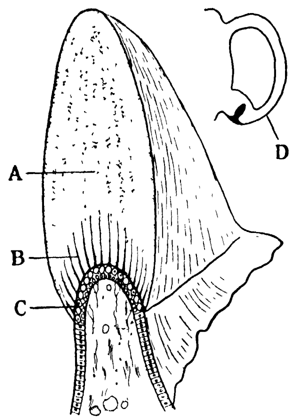


FIG. 64. CROSS SECTION OF CRISTA.
A = gelatinous mass (cupula). *B* = hair tuft. *C* = hair cell. *D* = diagram showing position of crista in ampulla of semicircular canals. Adapted from Schaffer

Receptors. The *semicircular canals*, three in each ear, lie approximately at right angles to one another in the three planes

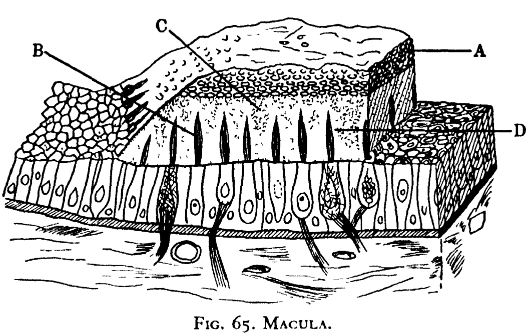


FIG. 65. MACULA.
A = otoliths. *B* = hair tuft. *C* = thick viscous substance *D* = lymph space. After Kolmer.

of space. At one end of every canal there is a swelling, an *ampulla*, containing a transverse membrane called the *crista*

(Fig. 64). This membrane is the functional organ—reflex, sensory or both, whichever it may be—of this part of the inner ear. It is a high, narrow, rounded prominence in the ampulla. The membrane contains hair cells, the hair tufts of which are imbedded in a gelatinous mass called the *cupula.*

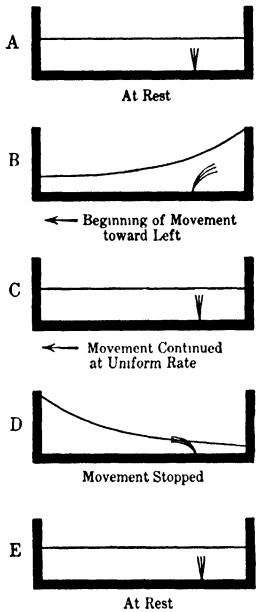


FIG. 66. DIAGRAM ILLUSTRATING THE HYDRODYNAMICAL PRINCIPLES OF THE SEMICIRCULAR CANALS.

Within each of the *vestibular sacs* (Fig. 63) there is a *macula* (Fig. 65), the functional organ of this part of the inner ear. The utricular macula occupies the one wall of the *utricle*, and its surface lies in the horizontal plane of the head. The saccular macula occupies a wall of the *saccule*, and its surface lies in the sagittal plane. The two maculae are thus at right angles to each other. The membrane of the macula is very similar to that of the crista. It also contains hair cells whose tufts project into a thick, viscous substance. The upper layer of this substance, extending beyond the ends of the hair tufts, contains a multitude of minute crystals of carbonate of lime that are known as the *otoliths.*

Theory. According to the generally accepted theory, the nerve endings in the ampullae are stimulated hydromechanically by the flow or pressure of the endolymph which fills the semicircular canals. When the head is turned, the endolymph, because of its inertia, does not immediately follow the walls of the canals. It lags behind, and the amount of the lag or the force exerted upon the cupula depends upon the rate of change in the speed of the stimulus. The physical principles involved are illustrated in Fig. 66, in which a pan, water and a tuft of hairs are substituted for a canal, endolymph and cupula. When

the system is at rest, as at *A*, the water and hair tuft are at equilibrium. If now the pan is moved to the left, as at *B*, the water will at first lag behind and bend the hair tuft toward the right. Then, if the movement is continuous and regular, it will gradually take up the movement and it and the hair tuft will return to an equilibrium, as at *C*. If a change is made in the rate of movement, or if the movement is suddenly interrupted, as at *D*, the water will rush forward, because of its

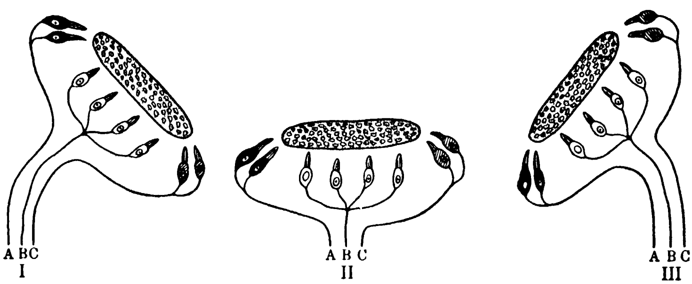


FIG. 67. MECHANICAL ACTIVATION OF THE MACULAE BY THE OTOLITHS.

inertia, and incline the tuft of hairs to the left. In either case, whether the rate is changed or the movement stopped, equilibrium will soon again be established, as at *E*.

Against this theory of functioning of the semicircular canals two objections have been raised. (1) The canals have lumens that are too small, and (2) the endolymph is too viscous for currents to be set up in it by rotation of the head. If these objections are sustained, we are left without any theory regarding the functioning of these organs.

The conventional view of vestibular stimulation is that the nerve endings in the maculae are activated mechanically by the otoliths. Whenever a change is made in the position of the head or the head is moved, the otoliths, because of their mass and inertia, stimulate the hair cells either (*a*) by pressure or (*b*) by pull. Fig. 67 shows how shifts in the position of the maculae will cause the weight of the otolithic membranes to

press (or pull) on different groups of hair cells. It is evident that any variation of the position of the head will be attended by the stimulation of different groups of hair cells.

Changes in the acceleration or direction of rectilinear movements of the head must also produce variations in the pressure (or pull) of the otoliths upon the hair cells. If, for example, a sudden movement is made toward the right (see Fig. 67, II), the otolith may not immediately take up the motion, but may lag behind and press against the hair cells to the left. Similar effects attend movements to the left, forward and backward, up and down and in the intermediate directions. Alteration in the direction of movement would thus stimulate different hair cells of the maculae.

SENSIBILITY OF THE INTERNAL ORGANS

The experiences from the internal organs and tissues of the abdominal and thoracic regions of the body are classified under the general term *organic sensations*. Those derived from the abdominal regions, excepting those of sex, are called more specifically *visceral*.

Our knowledge of the nature and origin of these experiences and of the part they play in our behavior and conscious life is meager, because of their vague and indefinite character, and because of the inaccessibility of the tissues involved. What we know of them comes from various sources—surgery, physiology, pathology, psychology—and much of it is conflicting as well as fragmentary and incomplete.

Evidence from surgery is the most precise—and probably the most misleading. When the abdominal or thoracic cavities are opened under local anesthetics, which affect only the skin and the underlying connective tissue, it is found that the internal organs—lungs, heart, esophagus, stomach, intestines, liver, kidney, etc.—are apparently insensitive. They may be burned, cooled, cut, crushed, torn, without yielding experience of any kind. The serous membranes—peritoneum, pericardium, pleura,

tunica vaginalis—and the serous and muscular layers of the diaphragm, however, yield pain; and the peritoneum and diaphragm seem also to respond with pressure.

The insensitivity of the internal organs appears from other points of view, however, to be an artifact. All the internal organs are abundantly supplied with afferent nerves. There is no *a priori* reason for questioning their sensitivity. Their failure to respond to such stimuli as the knife, the forceps, the cautery, may be due, not to their normal lack of sensitivity, but to the effects of the anesthetic, of the exposure or of the surgical shock, or to the ineffectiveness of external attack. Moreover, it seems reasonable to suppose that the stimuli most capable of exciting these afferent nerves should be the natural processes inherent in the organs.

SOMESTHETIC PERCEPTION

The complex experiences derived from the tissues and organs of the body may be divided into five general types or classes: (1) the touch blends; (2) the organic complexes; (3) spatial discrimination; (4) the static perceptions, and (5) the perception of movement. As the last three topics receive special consideration elsewhere (see pp. 239-241, 262f.), we shall consider here only the first two.

Touch blends. Our tactual experiences are normally derived from the simultaneous stimulation of a number of organs and tissues. An object placed upon the skin affects, not only the various cutaneous end-organs, but the subcutaneous as well. Indeed the separate stimulation of any end-organ or tissue is, as we have seen, a matter of great difficulty, and it rarely if ever occurs outside the safeguards of an experiment. Such experiences as wet, dry, hard, soft, smooth, rough, sharp, dull, oily, sticky, clammy, slimy, mushy, soggy, doughy, spongy, to name but a few of the more common, are the results of complex stimulation and the fusion, integration or blending together of many qualities.

Experimental studies have been made of a few of these

blends. It has been found that *wetness* is an integration of cutaneous pressure and cold in which the cold has the greater intensity and clearness. *Oiliness* is an integration of warmth and light pressure; and *burning heat* is a combination of heat and pain. *Stickiness* is a patterned experience of light contacts and pressures, whose intensity varies and shifts rapidly among the different points constituting the pattern. *Roughness,* similarly, is an areal pattern formed under movement from a fusion of cutaneous and subcutaneous pressure. *Clamminess* is more complex; it is an integration of cold and of softness (which is itself a blend) and of certain imaginal components.

Organic complexes. The organic complexes, such as hunger, thirst, nausea and the calls and performances of the wants of nature are fusions, blends or integrations of pain and pressure. These are all without qualitative uniqueness or novelty except as patterns. *Hunger* is a complex of muscular pain and pressure with a specific localization, the pit of the stomach. *Thirst* is a similar integration localized in the mouth. *Nausea* resembles hunger, and in the incipient stages may even be confused with it. *Stuffiness, suffocation* and *oppression* are similarly patterns of pressure and of pain localized in the thoracic cavities.

REFERENCES

I. **Cutaneous qualities:**

1. Dallenbach, K. M. The temperature spots and end-organs. *Amer. J. Psychol.,* 1927, 39, 402-427. (With Burnett, N. C.) The experience of heat. *Ibid.,* 1927, 38, 418-431. (With Stone, L. J.) Adaptation to the pain of radiant heat. *Ibid.,* 1934, 46, 229-242.
2. Luciani, L. Cutaneous sensibility. In *Human physiology.* London: Macmillan and Co., 1917. IV, pp. 1-55.
3. Nafe, J. P. The pressure, pain and temperature senses. In *Handbook of general experimental psychology.* Worcester, Mass.: Clark University Press, 1934. Pp. 1037-1087.
4. Sherrington, C. S. Cutaneous sensations. In E. A. Schäfer's *Textbook of physiology.* Edinburgh: Y. J. Pentland, 1900. II, pp. 920-1001.

II. Sensibility of the mucous membranes:

5. BORING, E. G. The sensations of the alimentary canal. *Amer. J. Psychol.*, 1915, 26, 1-57.
6. HERTZ, A. F., *The sensibility of the alimentary canal.* London: Oxford University Press, 1911.

III. Kinesthesis:

7. SHERRINGTON, C. S. The muscular sense. *Op. cit.* (4), pp. 1002-1025.
8. TITCHENER, E. B. *A textbook of psychology.* New York: Macmillan Co., 1909. Pp. 160-182.

IV. Somesthesis from the inner ear:

9. DUSSER DE BARENNE, J. G. The labyrinthine and postural mechanisms. In *Handbook of general experimental psychology.* Worcester, Mass.: Clark University Press, 1934. Pp. 204-246.
10. GRIFFITH, C. R. The organic effects of repeated bodily rotation. *J. Exper. Psychol.*, 1920, 3, 15-46. An experimental study of dizziness. *Ibid.*, 89-125.
11. PIKE, F. H. The function of the vestibular apparatus. *Physiol. Rev.*, 1923, 2, 209-239.

V. Sensibility of the internal organs:

12. BORING, E. G. Processes referred to the alimentary and urinary tracts. *Psychol Rev.*, 1915, 22, 306-331.
13. LUCIANI, L. Sensibility of the internal organs. *Op. cit.* (2), pp. 57-125.
14. TITCHENER, E. B. *Op. cit.* (8), pp. 183-193.

VI. Touch blends:

15. BENTLEY, M. The synthetic experiment. *Amer. J. Psychol.*, 1900, 11, 405-425.
16. FERRALL, S. C., and DALLENBACH, K. M. The analysis and synthesis of burning heat. *Ibid.*, 1930, 42, 72-82.
17. TUNG, S. The integration of punctiform cold and pressure. *Ibid.*, 1921, 32, 421-424.
18. ZIGLER, M. J. An experimental study of the perception of stickiness. *Ibid.*, 1923, 34, 73-84.

VII. Organic complexes:

19. BORING, E. G., and LUCE, A. The psychological basis of appetite. *Amer. J. Psychol.*, 1917, 28, 443-453.
20. CANNON, W. B. Hunger and thirst. In *Handbook of general experimental psychology.* Worcester, Mass.: Clark University Press, 1934. Pp. 247-263.
21. PACK, G. T. New experiments on the nature of the sensation of thirst. *Amer. J. Physiol.*, 1923, 65, 346-349.

Chapter 8

INTENSITY

Simple sensory experience varies in at least four different ways: in quality, in intensity, in its spatial relations and in its temporal relations. The four preceding chapters have dealt primarily with the problem of sensory quality and only incidentally with related problems of sensory intensity. The present chapter considers the general problem of intensity as it arises in all the senses.

INTENSITY AS A DIMENSION OF CONSCIOUSNESS

The difference between a high and a low tone in the musical scale is qualitative, but the difference between a loud tone and a soft tone of the same pitch is intensitive. It used to be supposed that every sensation could be defined in terms of these two attributes, quality and intensity, that the quality of a sensation establishes its character and gives it its name, and that intensity is the degree or quantity of the sensation. In the phrase *very bitter,* the adjective *bitter* denotes a quality and classifies the taste, but the adverb *very* pertains to the intensity *of* the bitter taste. This belief comes about because it reflects the common-sense view of the physics of the stimulus. Sensory quality is correlated with gross qualitative distinctions among stimuli, such as light, sound, heat, sugar, musk. Intensity, however, is varied by changing the strength of these stimuli: more sugar in solution gives a more intense sweet, more heat an intenser warmth.

Pure introspection would never have established the distinction between quality and intensity. All we should gain from

This chapter was written by Edwin G. Boring of Harvard University.

experience directly, if we had no knowledge of the stimuli that underlie it, would be the fact that it can change and that a particular experience can change in more than one way. There would be nothing about all qualitative changes to differentiate them from all intensitive changes. This fact is especially apparent in the case of visual intensity, which introspection fails to distinguish from visual quality. Let us consider the series of grays from black to white. If the series is given as surface-colors by a set of a hundred pieces of paper, we seem to have a qualitative series: no one gray is 'more' than another gray, and black is not a 'zero' but just as much a sensation as white. However, if the corresponding series of film-colors is given by gradually increasing the illumination from the zero of complete darkness to a maximum of brilliant white, then the changes appear to be intensitive. Thus the psychologist, in classifying the modes of sensory variation as qualitative or intensitive, must have some regard to the stimulus as well as to the conscious experience itself.

There is reason to believe that any sensory experience, whatever its quality, can vary intensitively. Most of the examples come readily to mind. Tones are soft or loud; noises, faint or loud; warmths and colds, mild or extreme; pressures, light or heavy; pains, mild or acute; tastes and odors, faint or strong. There is no doubt that vision likewise is intensitive, but it is difficult to define the dimension. The problem of the grays and the achromatic brightnesses has been mentioned in Chapter 4 (pp. 74f., 92f.), but the saturation of colors is also an intensitive dimension. We should undoubtedly have a satisfactory criterion of what to call intensitive if we knew more about the physiology of afferent nerve conduction.

THE STIMULUS TO INTENSITY

It is often said that energy is the stimulus to intensity. Speaking broadly this statement is correct. If the stimulus is increased in energy but remains otherwise unchanged, we generally find an increase in the sensory intensity but relatively

little change in the sensory quality. If a steel ball is dropped from a fixed height on an ebony plate, we get a noise of a given intensity, and, if the height of fall of the ball is increased, the energy of the fall and also the energy of the resultant sound are increased, so that the noise is louder. For the most part the change is intensitive, although the relationship turns out not to be simple.

However, it must be pointed out that our knowledge of sensation is still insufficient for us to be able to control the effective energy of all stimuli. Nowadays visual and auditory stimuli are best defined with respect to energy, but the other senses have not yet lent themselves to such refinement. Intensities of taste and smell are usually varied by changing the amount of the stimulus substance that can be effective. For taste, a solution of sugar (or quinine, or salt, or acid) is made more concentrated if the sensory intensity is to be increased. For smell, the concentration of an odorous liquid may be increased, or, as in the use of the olfactometer (see p. 151), the air to the nostrils may be drawn over a larger surface of an odorous substance. In neither taste nor smell, however, do we know enough of the manner of action of the stimulus to be able to say how energy is involved. In the somesthetic senses the stimulus to intensity is generally expressed in terms of force for pressure and pain or of temperature for warmth and cold. Thus cutaneous pressure is varied in degree by variation of force upon the skin, and cutaneous pain is varied intensitively by control of the force of a needle-point. In all the experimental research upon sensory discrimination for lifted weights the stimulus is measured in weight (force, grams) and not in terms of energy.

As a matter of fact the true nature of the stimulus to intensity has to be discovered by research. Von Frey made such a discovery when he demonstrated that the threshold for cutaneous pressure is constant when expressed as the total force on the skin divided by the diameter of the stimulating surface. (The law holds only when the stimulus is very small, like the end of a hair pressing against the skin.) In other words, von

Frey discovered that stimuli of different sizes and pressures are equivalent when their values expressed in grams per millimeter are the same. One might rather have expected that grams alone or else grams per square millimeter would be the proper stimulus, and von Frey's discovery not only told him something unpredictable about the stimulus, but also provided him with grounds for a theory of the physiological process in the receptor.

In experimental research on hearing, the control of the stimulus has been brought to a high degree of refinement, especially in the work with pure tones which are precisely defined when frequency and amplitude have been accurately determined. There are several units by which the energy of tones is measured, but they are all related in definite ways. Thus the stimulus to tonal intensity may be expressed as 'pressure amplitude' (dynes per square centimeter), or as power (microwatts), or as 'loudness' (decibels).[1] A very soft whisper uses about 0.001 microwatt of power, and a very loud shout may use 1000 microwatts, a million times as much power as the whisper. The intensitive difference between these two sounds is 60 decibels. Under many conditions a single decibel comes somewhere near the differential limen for intensitive discrimination.

Tones furnish an excellent example of the way in which sensory intensity is not in a simple manner dependent upon the energy of the stimulus, for both the intensity and the pitch of tones depend upon both the energy and frequency of stimulus. In Fig. 34 (p. 118) we can draw a line that connects tones of a given constant intensity, and we find that such a line approximately parallels the line of the absolute threshold. As the frequency is increased (up to about 2000 ~), the energy must be decreased if the sensed intensity is to remain

[1] A decibel is a relational unit. It specifies the power of one tone with respect to the power of another, and the number of decibels is equal to ten times the logarithm of the ratio of these powers. If Fechner's law were precisely correct as applied to tones (see pp. 200f.), then decibels would truly measure the subjective loudness of a sound. The decibel is sometimes called a unit of loudness because under some conditions Fechner's law is approximated. See p. 119.

constant; thereafter energy must be increased with further increases of frequency if intensity is still to be kept constant. If intensity depended upon energy alone, these lines would be straight horizontals. Nor are the lines for constant pitch exact verticals in such a diagram. When the energy of a frequency is increased, the resultant pitch changes slightly. In short, there is no simple correspondence between the dimensions of the stimulus and the dimensions of the sensation.

In the field of vision there has been as careful control of the energetic aspects of the stimulus as there has been for audition, and it becomes even more evident here how complex the relationships may be and how unlikely it is that any simple relation between the stimulus and the sensory intensity would ever express the precise truth. These complexities have already been expounded in Chapter 4; here we need only to note their significance in a new context. Light, the visual stimulus, is defined by the wave-length and energy of its various constituents. The radiant energy of light is measured in some unit of power (*e.g.*, ergs per second, microwatts). However, the consideration of the stimulus without a knowledge of the manner in which it acts gets us a very little way indeed. We have to take into account the nature of the eye and the retina to understand the stimulus to visual intensity.

In the first place we must note that we have to consider the sensitivity of the eye before we generalize about the visual stimulus to intensity. Ultra-violet or infra-red light, no matter what the energy, is never a visual stimulus, and the sensitivity curve (Fig. 16, p. 72) shows that a constant visual intensity requires the least energy at a middle wave-length. Stimulus values can be corrected by photometric equations in accordance with the differences that the sensitivity curve represents, so that different wave-lengths can be used as having equivalent values for intensity. Furthermore in accurate research the observer makes his judgments of visual phenomena through a pinhole that is smaller than the pupil of the eye, so that variations of the size of the pupil are eliminated; or else the experimenter corrects the stimulus values for pupillary size.

In a variety of ways, therefore, it has been found that an adequate knowledge of the visual stimulus depends upon a knowledge of the way in which the eye responds to stimulation, and that 'the true physiological stimulus' is the pattern of retinal excitation.

Not only does visual intensity depend in part upon wavelength, but also hue depends in part upon the energy of the light. As the energy of light increases, the hues of the red-yellow-green end of the spectrum tend to approach the hue of the 'invariable yellow' that corresponds to 575 mμ, and the hues of the green-blue-violet end to converge upon the hue of the 'invariable blue' that corresponds to 475mμ. (Pure green is the crucial point of division of the spectrum at 505 mμ.) This dependence of hue upon energy is analogous to the dependence of pitch upon energy.

Two important facts have emerged from our discussion of the stimulus to intensity. (1) We have seen that sensory intensity in the different sense departments is not dependent in any simple way upon some single aspect of the stimulus, like its energy. The intensitive relationships between sensation and stimulus are complex; they are the result of much research, and they come out differently in different cases.

(2) We have seen further that the sort of research that inquires into the nature of the stimulus to intensity is directed upon physiological sensory processes. Especially in the field of vision have psychologists used their knowledge of the eye to transform the stimulus from what it is known to be outside the organism to what they may reasonably infer it to be at the retina. The inevitable step is to pass from the psychophysics of intensity to its psychophysiology, and there our information can be somewhat extended by what we know about the nerve impulse in afferent nerves. To this matter the next section is devoted.

PSYCHOPHYSIOLOGY OF INTENSITY

The discovery of the all-or-none principle of nerve conduction raised a difficulty in respect of the physiology of sensory

intensity. Hitherto it had been supposed that an afferent nerve fiber, linking a receptor with the central nervous system, would function for a single quality with various degrees of excitation, thus giving rise to corresponding degrees of sensory intensity. When it became known that a single fiber responds only with one magnitude of excitation, a need for a special explanation of intensity arose.

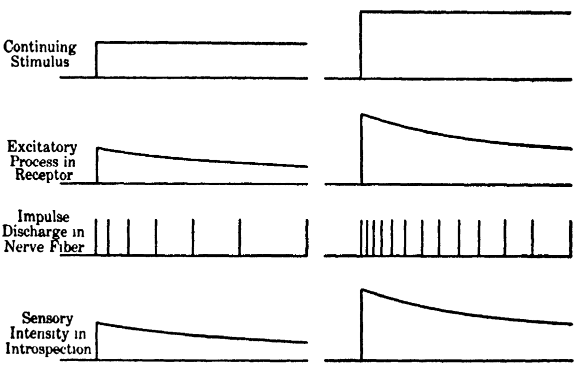


FIG. 68. FREQUENCY THEORY OF INTENSITY.

The figure shows the four successive phases in the stimulation of sensory intensity, for a weak stimulus at the left and a strong stimulus at the right. The excitatory process in the receptor is diminished by adaptation, and the diminution is reflected in the frequency of impulses and in the sensory intensity. Note that all nerve impulses are of equal strength under the all-or-none law, but that greater stimulus strength or sensation intensity means greater frequency in the nerve. Adapted from Adrian.

There are at present two accepted principles of the afferent conduction of sensory intensities. They are related to each other, and it is quite clear that neither of them is wholly incorrect.

1. The *frequency principle* of intensity grows out of recent discoveries by Adrian and other investigators associated with him (see pp. 18-21). It is found that a continued stimulus gives rise to a series of impulses along the excited fiber and that the rate of these impulses is increased when the stimulus is stronger (see Fig. 68). A brief stimulus also acts in somewhat the same way because its effect persists in the re-

ceptor long enough to set up a number of successive impulses. It is easy to see how this result comes about. The stimulus discharges the fiber. The fiber becomes refractory and the stimulus is ineffective. At some point in its relative refractory period the fiber recovers enough for the stimulus to discharge it again, and this process is repeated (see Fig. 4). If the stimulus is stronger, it discharges the fiber sooner (when the fiber has recovered less completely): thus the stronger stimulus produces a greater frequency of impulses. Although most of the observations that have led to this knowledge have been made upon nerves that contain many fibers, Adrian has shown that the relationship may hold for a single isolated fiber. It thus seems extremely probable that a greater frequency of impulses in a nerve fiber may lead to an increased sensory intensity.

2. The *multiple-fiber principle* of intensity is that the greater intensities result from the excitation of a larger number of fibers. It is a well-recognized fact that the excitation from small visual stimuli tends to spread in the retina when the strength of the stimulus is increased; and it seems reasonable that a large visual stimulus should also excite more retinal receptors as its strength increases, since the receptors are not likely to have the same thresholds and the intenser stimulus would pass the threshold of a larger number of receptors (see pp. 90f.). It is plain that when the stronger mechanical and thermal stimuli act on the skin or internally, their effects spread further and thus involve a greater number of fibers. There is also some reason to believe that intense tones activate more auditory fibers than do weak tones (see p. 138).

There is nothing incompatible between these two principles. Intensity may depend upon the amount of excitation. Excitation cannot vary at the instant of discharge in the single nerve fiber (all-or-none law), but the total excitation can vary with respect to the number of successive impulses in the single fiber or to the total number of fibers with impulses in them.

The *volley theory* combines these two principles in a particular manner. It assumes that, as stimulus strength increases,

the number of receptors brought into function also increases, that the frequency of impulses in any fiber depends on the strength of the stimulus, but that the frequencies in different fibers will differ for the same strength, being in part functions of the sensitivity of the fiber. This theory was developed by Wever and Bray in an attempt to explain the nature of impulses in the auditory nerve, and it has been already expounded in an earlier chapter (see pp. 136-139 and Fig. 42). It seems plain that we are here on the right track. Sensory intensity depends upon the total amount of peripheral excitation in some limited nerve tract. The total must involve many impulses in many fibers, or successive impulses in the same fiber, or both.

THE WEBER-FECHNER FUNCTION

It is fairly obvious that, when the strength of a stimulus is increased, the resultant sensory intensity follows a law of diminishing returns. If to two lighted candles in a room a third be added, there is a greater increase in the sensed illumination than there would be if a twenty-first candle were added to twenty lighted candles. The additional light has more effect when added to a lesser illumination. This relationship between sensory intensity and the strength of its stimulus is called the Weber-Fechner function because it has been expressed, although inexactly, in two historically famous laws: Weber's law and Fechner's law.

A century ago the physiologist, E. H. Weber, laid it down as a general principle that the increment of stimulus necessary to produce a just noticeable increase in the intensity of sensation is a constant fraction of the magnitude of the stimulus; that is to say, if R is the strength of the stimulus, and if ΔR is the increment of it that is required to produce a just noticeable difference, then

$$\Delta R/R = \text{a constant.} \quad \ldots \quad \text{Weber's law}$$

The ratio, $\Delta R/R$, has been called the Weber fraction (or, incorrectly, the Fechner fraction). The solid line of Fig. 69, *A*

depicts Weber's law graphically, since it shows the Weber fraction as constant for all intensities of stimulus. A simpler way of stating this same relationship is to say that the just

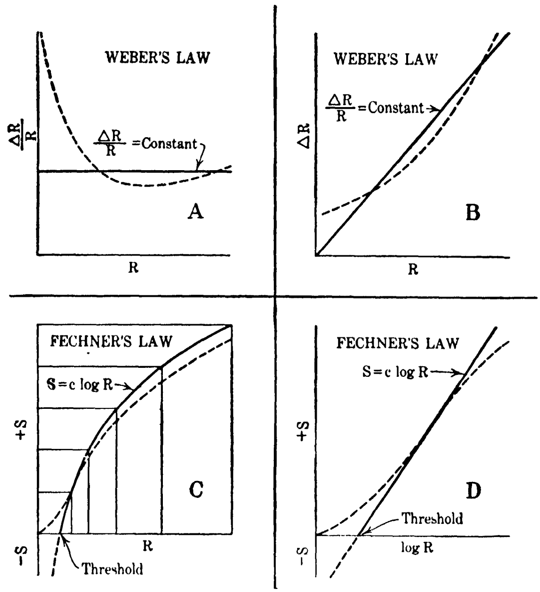


FIG. 69. WEBER-FECHNER FUNCTION.

R = total strength of stimulus. ΔR = just noticeable difference in the stimulus, actually the differential threshold. S = intensity of sensation, plotted as the number of ΔR's above the threshold and thus assuming that all ΔR's are *subjectively* equal. *A, B:* two ways of plotting Weber's law. The test of Weber's law is that the data in *A* should lie in a horizontal straight line, or that the data in *B* should lie in a straight line that passes through the origin. *C, D:* two ways of plotting Fechner's law. *C* shows how Fechner's law assumes that a geometric series of R's corresponds to an arithmetic series of S's. The test for Fechner's law is to note whether the data follow a straight line, as in *D*. The theoretical form of Fechner's law implies the existence of negative sensation, $-S$. In all four diagrams the dotted curves show the most usual ways in which the theoretical forms deviate from actually obtained data. In *A*, $\Delta R/R$ is fairly constant in middle ranges, but increases somewhat at great intensities and very much indeed at low intensities. In *D* this deviation makes the curve S-shaped, although it approximates a straight line in the middle portion.

noticeable difference, ΔR, increases proportionately with R, as is shown by the solid line of Fig. 69, *B*.

However, Weber's law seems never to be realized in practice for a wide range of intensities. The Weber fraction is apt to

follow a function like the dotted line of Fig. 69, *A*. Such a curve is definitely established as correct for visual intensity (see Fig. 70), and a similar function without the rise at the greater intensities holds for tone (see Fig. 71). It seems therefore to be a general rule that the Weber fraction is large at the absolute threshold, that it decreases, at first rapidly and then slowly, toward the middle intensities, and that the horizontal part of the curve gives the impression of approximate constancy in this middle region. These statements are true for all sensory intensitive functions that have been thoroughly worked out. They hold, for example, in the seven cases listed in Table XI. Beyond the middle range of intensities the Weber fraction may increase again (see Fig. 70) as it does for the visual, the gustatory and the three somesthetic cases of Table XI, or it may reach a minimum and not rise (see Fig. 71) as for tone and smell. Of course, one cannot say for tone whether a rise might be found if the function could be carried to still greater intensities. There is a plausibility about supposing that a sensory mechanism is especially well adapted to some normal mid-intensity, and that sensitivity should be less at both the lower and upper intensitive extremes. (See also Fig. 48, p. 152, for smell.)

Although there is no single value of the Weber fraction that holds for all intensities of any single sense, it is nevertheless possible to determine the minimal Weber fraction for each *sense*. These values are given in Table XI for seven cases. The fractions there are accurately established under the conditions of the experiments and methods that yielded them, but they depend in part upon chance factors. For instance, ΔR in these experiments is not strictly the 'just noticeable difference.' It is a statistically determined differential threshold or else some approximately comparable measure of sensitivity. Thus in less thorough researches the Weber fraction for vision has been observed as 1/100, and even once as 1/195. Nevertheless the gross differences are clear. Tonal intensitive sensitivity is less than visual. The skin is not so sensitive to difference of

pressure as is the kinesthetic lifting mechanism to difference of weight.

TABLE XI

Minimal Weber Fractions

For all cases below, except tones and smells, the Weber fraction has a minimal value in the middle of the range of intensities, and sensitivity is less at both extremes, as in Fig. 70. The minimal values given for tone and smell are for maximal intensities after the Weber function has leveled off, as in Fig. 71. None of these values must be taken too literally, since the determination of ΔR in the Weber fraction, $\Delta R/R$, depends in part upon various particulars of the psychophysical method used to measure it. Nevertheless the difference between the Weber fractions 1/77 and 1/5 is great; there can be no doubt that different sensory mechanisms differ significantly in sensitivity.

	Weber ratio	Weber fraction
Deep pressure, from the skin and subcutaneous tissue. At and near 400 gm. Biedermann and Löwit (1875)	0 013	1/77
Visual brightness. At and near 1000 photons. Computed from data of König and Brodhun (1889). Accurate data from Aubert (1865) and Blanchard (1918) agree closely, though the fraction has often been set at 1/100 and lower.	0.016	1/60
Lifted weights (kinesthesis). At about 300 gm. Oberlin (1934).	0.019	1/52
Tone. For 1000 cycles, at and about 100 db. above the absolute threshold. Knudsen (1923) . . .	0 088	1/12
Smell. For rubber, at 200 olfacties. Zigler and Holway (1934)	0 104	1/11
Cutaneous pressure, on an isolated spot. At and near 5 gm. per mm. Gatti and Dodge (1929).	0 136	1/7
Taste. For saline solution, at about 3 moles per liter concentration. Holway and Hurvich (1934)	0 200	1/5

In a famous book on psychophysics published in 1860, G. T. Fechner extended Weber's law so as to make it apply to the general relation between sensory intensity and the strength of

the stimulus. Fechner argued that all just noticeable differences in sensation, being 'just noticeable,' must be equal, and that, since equal just noticeable differences are caused by a change in the stimulus that is a constant fraction of its amount, other larger equal differences must also depend on constant fractional

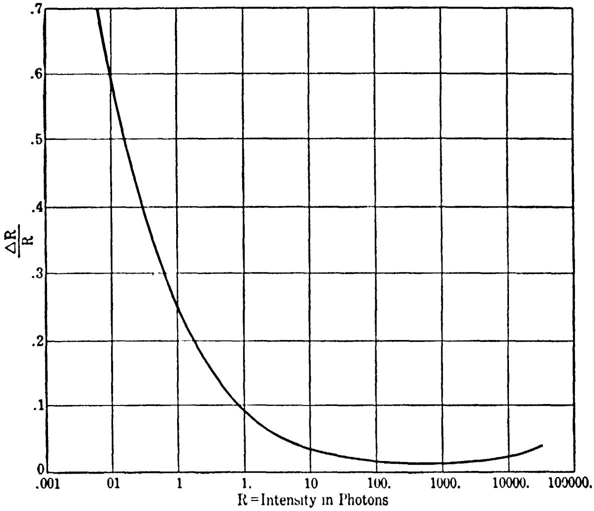


FIG. 70. WEBER FUNCTION FOR VISUAL INTENSITY.

R = retinal illumination in photons of white light ΔR = average differential threshold for the eyes of Konig and Brodhun If the Weber fraction, $\Delta R/R$, were constant, the function would be a straight horizontal line. Actually Weber's law is approximated between 10 and 1000 photons Maximal sensitivity ($\Delta R/R$ = 0.0165) is at about 1000 photons. The sensitivity decreases for great as well as for low intensities. The photons have been scaled off logarithmically in order to make the abscissa scale closer to a scale of subjectively equal units, but this displacement does not affect the test of Weber's law, which requires constancy of $\Delta R/R$ for any scale of intensities.

changes in the stimulus. Fechner put this general law into an equation. If R is the strength of the stimulus measured in units of the absolute threshold, if S is the resultant intensity of sensation and if c is some determinable constant, then:

$$S = c \log R \ldots\ldots \text{Fechner's law}$$

The curve of this function, shown in Fig. 69, C, is drawn to show that equal increases in sensation are produced by those

increases in the stimulus which have a constant ratio to the total stimulus magnitude. Another way of describing this logarithmic relationship is to say that an arithmetic series of sensory intensities is dependent, point for point, upon a geometric series of stimuli. While Fig. 69, *C* shows clearly the nature of Fechner's function, Fig. 69, *D* is the form by which its accuracy can best be tested. When *S* is plotted against log

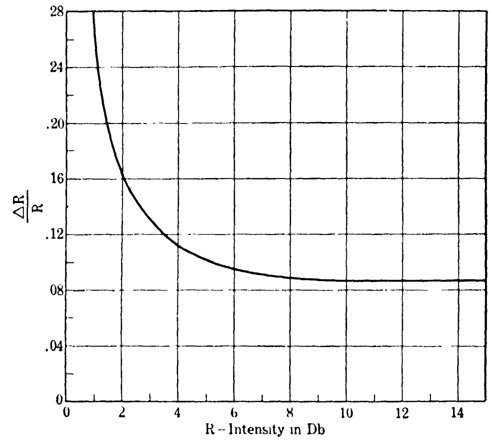


FIG. 71. WEBER FUNCTION FOR TONAL INTENSITY.

Knudsen's data for a tone of 1000 cycles. R = intensity of tone in decibels (db). ΔR = differential threshold The Weber fraction is constant and minimal ($\Delta R/R$ = 0.088) above 10 db., and practically so above 6 db. This curve does not rise at the greatest intensities (see Fig. 70), but it might if it could be continued further without injury to the ear. The decibels of the abscissa scale are measured from the absolute threshold, and are a logarithmic scale (see the photon scale of Fig 70) since a decibel here is 1/10 of the logarithm of the ratio of the power of the tone in question to the power of liminal tone at the threshold.

R (as in Fig. 69, *D*), observed data must lie in a straight line if Fechner's law is verified.

Since Fechner's law depends upon Weber's law, its exact verification is not to be expected. The Fechner function is S-shaped, like the dotted line of Fig. 69, *D,* if the Weber function is U-shaped, like the dotted line of Fig. 69, *A*. Hoagland (6) has argued that all Fechner functions are likely to be S-shaped, although it is plain that the function for tone, which we have just discussed, would lack the upper shoulder of the S.

In general, however, it is best to limit consideration to Weber's law in the form of Fig. 69, *A*.

On the other hand, it must be noted that a curvilinear function of the general form of Fig. 69, *C* is presupposed by common sense in a multitude of practical judgments. It is the general law of the psychical relativity of magnitudes. A piece of gold may mean life to a beggar but nothing noticeable as an addition to the treasury of Croesus. An inch added to one's stature seems less important than an inch added to one's nose.

INTENSITIVE SENSITIVITY

Under optimal conditions the energy required to excite sensation is surprisingly small. The absolute intensitive thresholds (the liminal values for the just noticeable intensity) are quite low as compared with most of the stimulus magnitudes that the human organism must constantly meet and take into account.

The threshold for visual brightness in the dark-adapted central field of vision is about 0.275 microwatt per sq. m. of illuminated surface. (A microwatt is a millionth of a watt; the watt, the unit in which electric-light bulbs are rated, represents the expenditure of 10 million ergs of energy per second; an erg is approximately the energy required to raise a gram weight 0.001 cm. against gravity.) The threshold seems small when expressed in this way. However, it is only about one-thousandth of this value on the peripheral retina. These figures apply to large stimuli.

Very small stimuli, like the stars, exhibit retinal sensitivity even more strikingly. The power of a faint star, visible in peripheral vision, is about 10^{-8} microwatt (100 million millionths of a watt; one thousand millionths of an erg per second). Such a star would have to shine on the retina for forty years to deliver an erg of energy to it. If it can be perceived in a fifth of a second, as seems probable, then the retina is some 30,000 times as sensitive as the most sensitive radiometer which

The intensitive threshold for tone varies greatly with the pitch of the tone. It is least for the frequencies in the region of 2000~ to 4000~, where it is approximately 4×10^{9} microwatt. This value is about half as large as the figures given for visual brightness in the preceding paragraph. It is generally supposed that the retina can be excited by less energy than can excite the ear, but the safest generalization is that the sensibilities of the eye and the ear are, for optimal conditions, of the same order. Certain investigators have computed that, if the minimal auditory threshold were only 6 db. lower, the ear could detect as sound the movement of the molecules of the air.

The energy required for the threshold increases tremendously toward the limits of hearing. At 20~ the energy is about six million times as great as at 3000~; at 20,000~ the energy is about six billion times as great. Beyond the 'limits of hearing' no energy that it is safe to deliver to the auditory organ will elicit tone.

We can make no such precise statement about the somesthetic thresholds because there has not been in this field the same accurate control and measurement of the stimulus. For large areas of the skin of the hand the limens for both cold and warmth, measured from physiological zero, have been found to be as low as 0.03° C. The minimal value for the range of thermal indifference from the cold limen to the warmth limen was, in this research, 0.071° C. There has been no attempt to translate such temperatures into statements of the transfer of energy.

In a thorough investigation of the pressure sense, in which the ends of hairs were used as stimuli, von Frey found the threshold to remain constant at about 0.83 gm. per mm., a unit which is got by dividing the pressure by the radius of the stimulus hair. For hairs ranging from 0.02 to 0.12 mm. in radius, the value of the limen in grams varied from 0.02 to 0.10 gm.; in areal pressure it varied from 15 to 2 gm. per sq. mm.; but expressed in grams per millimeter it remained constant. It has been estimated that the minimal stimulus

to pressure involves about 0.0001 erg, a value which is about 10,000 times as large as the limens for light and tone.

The threshold for cutaneous pain is even less accurately established because the essential nature of the algesic stimulus is not known. Thermal stimuli elicit pain about 20° C. above or below physiological zero. The pressure at which a needle-point or thorn elicits pain depends upon the conformation of the point. Any stimulus that is adequate to cutaneous pressure will at a greater strength bring out the algesic quality. Such thresholds for pain are from three to a hundred times as great as the corresponding thresholds for touch.

Olfactory sensitivity is great. There is no way, however, of expressing the strength of the stimulus in terms of energy; hence this sense cannot be compared in this regard with vision and audition. The putrid-smelling mercaptan creates a vile stench when the dilution of the stimulus is one in three hundred million parts of air, and the threshold is represented by a dilution of approximately one part in fifty thousand million. A single breath of this liminal stimulus would contain about 0.000,000,000,002,2 gm. of mercaptan, which is less than one three hundredth of the smallest amount of matter (sodium) that can be detected by the spectroscope. Still we are not anywhere near the limits of subdivision of a compound; this breath should contain about twenty-one thousand million molecules of mercaptan, a very great many of which would come into contact with the olfactory membrane.

The thresholds for taste are expressed in no such minute quantities of substance. The intensitive limen for sweet, when saccharose (sugar) is the stimulus, is 0.005 gm. per cc. of water, and, when saccharine is the stimulus, 0.00001 gm. per cc. The limen for bitter, when quinine is the stimulus, is 0.0000005 gm. per cc. Bitter seems to be the quality that is most often aroused by a very small concentration of the stimulus, but there is no general rule. (See pp. 144-146.)

Differential sensitivity does not involve as small quantities as we have been discussing for absolute sensitivity because of the operation of the principle of diminishing returns in ac-

cordance with the Weber-Fechner function. For instance, the threshold for tone at 2048~ is about 0.000,000,000,4 microwatt. The upper limit of intensity at this frequency, the point at which the tone becomes tactually uncomfortable, lies at about 5000 microwatts. If there were no such thing as the Weber-Fechner relationship and if the differential limens were all steps equal in size to the stimulus limen, then it would be possible, at this pitch, to discriminate about twelve million million different intensities. Actually about 366 are discriminable. The number of intensities discriminable falls off to about a hundred at 125~ and 8000~ and becomes very small indeed at the limits of audible frequency.

Ordinary speech does not approach either of the intensitive extremes. A very soft whisper has a power of about 0.001 microwatt. A faint voice may drop to 0.1 microwatt without becoming a whisper. A very loud shout would be about 1000 microwatts. The power involved in ordinary speech is about 10 microwatts, and the air particles then move only about 0.01 mm. It would require a million persons all talking at once to produce the power necessary to light a small electric lamp. Factory and city noises may be much louder than speech, so loud that a shout cannot be heard. One noise survey found Thirty-fourth Street and Sixth Avenue the noisiest place in New York City. The power of that noise was 100,000 microwatts.

The best figures on the differential limens for visual brightness show 572 steps from no illumination at all up to maximal brilliance. About 30 of these steps are supposed to be furnished by the retinal rods, and the remainder by the cones. Again we see how the Weber-Fechner function reduces the number of discriminable differences, for the illumination of the disk of the sun is about sixty million million times the liminal illumination for a disk just perceptible in the dark. White paper in direct sunlight is only about sixty thousand million times the limen. These very great brightnesses injure the eyes. The stimuli for ordinary good vision lie between two million and two hundred million times the absolute threshold.

There is not much to say about differential sensitivity in the other senses. Taste and smell have not yielded readily to quantitative research. Gross tactual discrimination varies with the region of the skin, with the pressures used and consequently with the amount of involvement of the subcutaneous tissue; but within a fairly wide range of pressures the Weber fraction is fairly constant near 1/5 for an isolated cutaneous spot. The thermal differential limen for the skin seems to be about 0.3° C. Articular discrimination is good, and a movement of the elbow joint of less than one degree can be detected. In lifting weights the Weber fraction is about 1/40 at 100 gm. and tends to decrease as the stimulus weight increases; but one method makes it appear about constant between 50 and 500 gm.

REFERENCES

1. ADRIAN, E. D. *The basis of sensation.* New York: W. W. Norton and Co., 1928.
2. ADRIAN, E. D. *The mechanism of nervous action.* Philadelphia: University of Pennsylvania Press, 1932. Chapters 1 and 2.
3. BORING, E. G. *The physical dimensions of consciousness.* New York: D. Appleton-Century Co., 1933. Chapter 3.
4. FLETCHER, H. *Speech and hearing.* New York: D. Van Nostrand Co., 1929. Part I, Chapter 3; Part III, Chapters 2 and 3.
5. *Handbook of general experimental psychology.* Worcester, Mass.: Clark University Press, 1934. Chapters 3, and 13 to 20.
6. HOAGLAND, H. The Weber-Fechner law and the all-or-none theory. *J. Gen. Psychol.*, 1930, 3, 351-373.
7. TROLAND, L. T. *The principles of psychophysiology.* New York: D. Van Nostrand Co., 1930. II, especially sections 53, 54 and 61.

CHAPTER 9

THE PERCEPTION OF SPATIAL RELATIONS

The familiar objects about us belong to a spatial world, and we perceive them as having length, breadth and, in many cases, depth. Furthermore, these objects are localized at certain places: they may be on the surface of or within our own bodies; or they may be at some distance and in a certain direction away from ourselves. The localization of a particular object always involves a reference to some other object whose position we also perceive or already know. We perceive a fly on the hand, a pain in the heart, a tree beyond the wall, a star above us. Localization always involves a reference to the position of our own bodies; the meaning of the terms *up, down, below, above, right* and *left, near* and *far* changes as the position or posture of our bodies changes. Consequently we must also perceive the position of our own bodies relative to other objects. In general, to perceive an object at all is to perceive it at some place, *i.e.,* the localization of an object is bound up with the perception of the object itself, but in this chapter we shall be concerned solely with those aspects of our perception which are spatial.

TACTUAL SPACE PERCEPTION

Localization of points on the skin. Cutaneous sensations are localized with respect to the part of the skin stimulated. A fly alighting on the forehead is not confused with one on the neck; a pinch on a finger is distinguishable from one on the back of the hand. To the uncritical individual this fact is so obvious that it seems hardly worthy of mention, but the student of psychology will wonder how it can be possible that

This chapter was written by Warner Brown of the University of California.

sensations which are qualitatively alike in every other way can still be distinguished in terms of the particular part of the skin from which they arise. The fact is obvious; but the actual functional relations involved remain an outstanding puzzle which affords a good starting point for examining the problems of space perception in general. These problems have aroused the curiosity of keen thinkers since the time of Descartes and Berkeley, and they still remain highly significant as we slowly approach their solution through modern experimental methods. The human or animal body is a solid object, like any other physical body, possessing length, breadth and thickness. When we localize a point on or within our bodies, we demonstrate our ability to respond in terms of these three dimensions of space; in other words, we are perceiving space with respect to at least one physical object.

In making such responses we may describe the spot stimulated, verbally or by drawing, or by pointing to a prepared diagram; or we may move some member of the body so as to point toward, or touch, the spot. The latter method has been preferred in experimental work because, in this way, the errors of localization can be stated directly in measured units of distance and direction. In such work the eyes are usually kept covered while the stimulus is being applied. This permits the experimenter to mark definite points on the skin to be stimulated and prevents the subject from being influenced by irrelevant factors. Blind adults do not do better than seeing adults who do not use their eyes. In any event, the errors are never very large in terms of the width or length of the body or its limbs. On the back of the hand the errors are of the order of 5 mm. or less. Errors are smaller on the face, hands and feet than on other, less mobile or less frequently exposed, parts of the body. If the arms are crossed, or the fingers crossed, or the lips displaced from their usual resting position, systematic errors of localization occur.

The threshold of duality. When two points on the skin are simultaneously stimulated, they will feel like a single point

unless they are separated by a sufficient distance. The liminal value of this separation is known as the 'two-point threshold.' It ranges from a fraction of a millimeter on the finger tips, lips and eyelids to several centimeters on those parts of the body (such as the middle of the back) less used in active touching or less abundantly provided with cutaneous sense organs. If the two stimulating points are not set down quite simultaneously or if the pressure is not equal upon the two points, they are more likely to be felt as two separate qualities. Measurements made transversely (across a limb) are usually smaller and less variable than those made longitudinally (lengthwise). If pressure spots are stimulated, the values obtained for the limen are much reduced.

There is a tendency on the part of nearly all observers in two-point threshold experiments to give some two-point reports when only one point has been set down. Consequently it is customary in experimental procedure to employ stimulation by a single point as a frequent check.

Observers tend also to fall into a 'stimulus error,' in which, instead of reporting the experience of two clearly separated points, they assume a detective attitude and report as 'two' all those cases in which the stimulus is *inferred* to be a double point. Such an inference is possible because two points applied close together differ from one point, in that they feel blunt, dull, broad, like a line, circle, ellipse or some other two-dimensional figure. Analysis of these experiences affords cues upon which correct inferences *as to the stimulus* can be based even when there is no *direct awareness* of two separate pressures.

The values of the two-point threshold at different parts of the body are generally regarded as measures of cutaneous sensitivity. The tips of the fingers, for example, are more sensitive than the back of the hand, and differences between the two-point threshold at the two places may serve as measures of the difference between them in sensitivity. The values themselves, like all other threshold values, are not only statistical in nature (see p. 38), but also dependent upon the success with which

the experimenter has brought the many sources of error under control. Table XII gives values representative of the relative magnitudes of two-point thresholds which may be expected.

TABLE XII

REPRESENTATIVE VALUES OF THE TWO-POINT THRESHOLD

	Mm.
Tip of the tongue	1
Volar side of the last phalanx of the finger	2
Red part of the lips	5
Volar side of the second and dorsal side of the third phalanx of the finger	7
White of the lips, and metacarpus of the thumb	9
Cheek, and plantar side of the last phalanx of the great toe	11
Dorsal side of the first phalanx of the finger	16
Skin on the back part of cheek-bone, and forehead	23
Back of the hand	31
Kneepan, and surrounding region	36
Forearm, lower leg	40
Back of foot, neck, chest	54
Middle of the back, and of the upper arm and leg	68

Size and form of touched objects. The perceived distance of points on the skin above the two-point threshold is determined in part by the size of the two-point threshold at that place. For example, when two compass points a fixed distance apart are drawn along the arm from the elbow to the wrist and on down to the fingers, their separation appears to change, and a diagram of their perceived path drawn by the subject shows that they seem close together at regions where the threshold is large and far apart where the threshold is small.

As early as the time of Aristotle it had been observed that, if a marble is placed between the crossed index and middle finger, it will be felt as two marbles. In the accustomed position of the fingers (uncrossed) the two cutaneous areas would normally be stimulated by two different objects, and two objects would in consequence be perceived. This perception of duality

is so stable that it persists even when the fingers are crossed and the two areas stimulated by a single object (see Fig. 72).

The length of the edge of a card or similar linear distance is perceived as a continuous line whose length is determined with reference to the localizations of the points on the skin that are stimulated. The *length* perceived with a stimulus of this kind (filled space) is not the same as that perceived with a stimulus of two points (empty space) separated by a distance equal to the length of the card. In these cases the stimulus is brought into contact with the resting skin (passive touch).

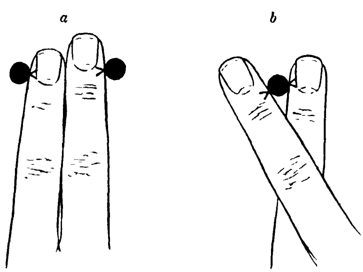


FIG. 72. ARISTOTLE'S ILLUSION

When the hand is moved across a stationary object (active touch), perceptions of length or distance are more accurate.

By passive touch the form of solid objects is apprehended imperfectly, and a simple geometrical solid seems very irregular; each of the several faces is much distorted, and the relations between them are difficult to grasp. By active touch the shapes of objects are much more distinctly perceived.

Touch and kinesthesis enable us with eyes closed to perceive the relative positions of the various members of the body at any particular moment, and we are able to locate, by moving one hand, the position of the other hand or of any of the other members with considerable precision.

Theory of tactual space perception. From the facts set forth above, two fundamental problems emerge. The first is that of spatial continuity. The receptors in the skin are arranged in a mosaic. Nevertheless when two or more punctiform

stimuli whose distances apart are less than the two-point limen are simultaneously applied to the skin, and also when the stimuli are linear or areal in form (*e.g.*, the edge or surface of a card), the resulting perceptions are linear or areal in pattern. Under these conditions the simultaneous stimulation of discrete receptors gives rise to a perception of a continuous surface. Since similar phenomena occur in visual space perception and also in the perception of movement when the stimuli are applied successively, its further discussion is deferred (see pp. 235, 261f.).

The second problem is that of cutaneous localization. The ability of animals to make appropriate movements of localization as in scratching an irritated spot is well known. Furthermore, when in an animal the brain is functionally separated from the spinal cord and the animal is then stimulated, the scratch reflex still occurs. When again the animal is prevented from executing the scratch reflex with the appropriate limb, the movement takes place with the opposite limb. There would seem to be a nervous connection between the receptors in the various parts of the skin and the motor systems involved in reaching the same areas. We might suppose, therefore, that cutaneous localization in some way involves the integration of the cutaneous quality and kinesthesis. The difficulty with this view is that the human individual is able to localize without movement. It has, therefore, been assumed that every sensation from the skin possesses a *local sign* indicating its point of origin. Some psychologists have thought that the local sign is of a qualitative or intensitive kind; others, that it may be an integration of qualities and intensities either from the skin alone or from the subcutaneous tissue also. Experiments have shown, however, that although differences in quality and intensity are found they have no uniform relation to differences in place. Another view is that the local sign is unconscious, that, although the sensory experiences from two cutaneous regions may be identical, the stimulations in the two cases, being actually different, arouse by association different localizing movements or different localizing visual imagery.

Despite the fact that localization is immediate even in the school child, there is considerable evidence that it had first to be learned, *i.e.,* the perception is a pattern the integration of which required practice. The infant goes through a long series of ill-directed motions before bringing his hand to his mouth or, later, in executing any coordinated movement. The blind man with his cane 'feels' the ground upon which he ventures. The automobile driver 'feels' the curb against which the wheel of his machine strikes. The surgeon 'feels' objects with his probe. We 'feel' through shoes or gloves. In many such instances the perception of spatial relations is just as immediate and unanalyzable as when the skin itself is in contact with the object. Since it is obvious that such perceptions are not referable to any innate structure already adapted to the work we must assume that they have been learned. Also, evidence of a pathological kind supports this view. Persons who have lost a limb by amputation still refer pains to the missing limb. There is a well-attested case of a man who had suffered a gunshot wound in the brain and who, when touched upon any part of his body, was unable to say where he was touched so long as he remained perfectly still. When, however, he was allowed to make exploratory movements he was eventually able to localize the impression. It would seem clear from this evidence that an adequate theory of cutaneous localization must be associative and in terms of the brain. Our knowledge of brain function, however, is still so fragmentary that a useful hypothesis cannot now be made.

VISUAL SPACE PERCEPTION

Visual perception of size. We know from common experience that there is a relation between the size of the stimulus and the size of the perceived object. Large things look large and small things look small. But we also know that large things when far away may look small. The perceived size of an object must, therefore, depend both upon the absolute magnitude of the stimulus and its distance from the ob-

server. The simplest way to deal quantitatively with these two variables is in terms of the visual angle subtended by the stimulus or by the retinal image (*i.e.,* the image of the stimulus which would fall on the retina; see Fig. 73).

The law of the visual angle is that the area of an object decreases with the square of the distance. This relation between the size of the retinal image and that of the perceived object is not constant, however. The retinal image of a coin held at ten inches from the eye is four times as large as when held at twenty inches, but for perception there is little if any difference in size. When familiar objects are close at hand there is, then, a tendency for their perceived size to remain

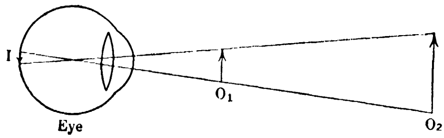


FIG. 73. THE VISUAL ANGLE.

constant even when seen at different distances. It is called the rule of *size constancy.* This rule does not hold, however, for objects at a remote distance. The apparent size of an object is also often related to surrounding objects. A tall man looks taller by the side of a short man. In Rostand's play the man-sized chanticleer is, on the stage, made to approximate the familiar size of a cock by increasing the size of the surrounding barnyard scenery; in puppet shows the opposite effect is reached by decreasing the size of familiar objects in the background.

The smallest object that can be seen (the minimum visible) depends upon several variables. A white spot on a black ground in sunlight can just be seen when the angle subtended is about 10 or 12 sec.; a black spot on a white ground in diffuse daylight requires an angle of 25-36 sec. of arc. Thus the brightness of the stimulus, its relation to its background and the intensity of the illumination are all important factors. In addition to these, the relative sensitivity of the retina which, it will

be recalled, depends upon the degree of adaptation, must also be taken into account.

The threshold of duality and the perception of details. When two stimuli simultaneously excite the retina, two objects will be perceived if the distance between the stimuli is sufficiently great. The separation that will just permit the perception of two objects is usually measured by the visual angle. The result corresponds to the two-point limen on the skin. Its value, like the minimum visible, depends upon the size and brightness of the stimuli, contrast with the background, and the intensity of the illumination. In a well-lighted room two black dots or lines on a white background can just be seen as two when they subtend an angle of 60 sec. of arc. This value will be increased with very high illuminations, as, for example, when the sun shines in the eyes, or with low illuminations, as at twilight. It will be lowered with brighter stimuli and with greater contrast between stimuli and background. These values presume, however, that the stimuli fall on or near the fovea of the retina and also that the eye as an optical instrument is normal. The limen is smaller at the fovea than at the periphery of the retina. For this reason the fovea is said to be the place of clearest vision and when we desire to examine an object closely we look directly at it, so that the image of the object falls on the fovea. If the eye is not normal the images of the two stimuli either will not be properly focused on the retina or there may be distortions of various kinds.

The limen of visual duality is regarded as the measure of visual acuity, the ability of the eye to distinguish small differences in objects. The oculist takes as his standard for normal vision the angle of 60 sec.; if the eye can see a letter the distance between whose parts, *e.g.*, the horizontal bars of an E, subtend an angle of 60 sec., it is in that respect considered normal. If, on the other hand, the angle subtended by the parts of the letter must be, say, 120 sec., before the letter can be distinctly seen, visual acuity is said to be one-half normal. The test letters are, of course, shown under good illumination.

General conditions of the perception of distance. Traditionally, the distance of an object away from the observer is referred to as the 'third dimension' in contradistinction to the first two dimensions 'width and length' which are involved in the mere extent of the object or of its image. Perception of this third dimension involves several factors some of which are more or less independent and some of which are interrelated. The factors that are discussed in the following paragraphs operate equally well in monocular and binocular vision.

The first is the *interposition of objects.* Objects in the foreground of the visual field are seen in their entirety. Objects in

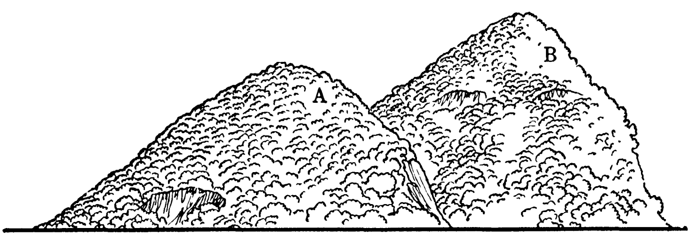


FIG. 74. INTERPOSITION.

the background are likely to be obscured in part by the intervening objects. We therefore perceive objects which are complete in outline as placed in front of similar objects which are incomplete. This perception is particularly likely to result when the object seen as entire is resting on the ground, *i.e.,* if the pattern of the plane upon which the objects stand is uninterrupted as far as the bottom of one object. Completeness of configuration does not, as a rule, afford an adequate index of the distance between objects, but it establishes an order of *relative* distance in that the object in the foreground is perceived to be *in front of* the one in the background, and this fundamental relation when once established determines the *direction* of a difference in distance, the actual *amount* of which may be determined with the help of other factors.

A second factor in the perception of the distance of ob-

jects is *linear perspective.* Whenever we see an object as much smaller than we know it to be, we perceive it as at a distance. If a number of objects of the same size were distributed in front of and at different distances from the observer and their sides connected with each other by lines, the lines would, as the objects became smaller and smaller, converge in the remote distance. In the same way when we look down a highway parallel lines are perceived as converging to a point. If these lines are crossed by other lines, angles are formed in accordance with simple geometric principles. Thus the right angles which are formed by the wires and cross-arms of the telegraph line

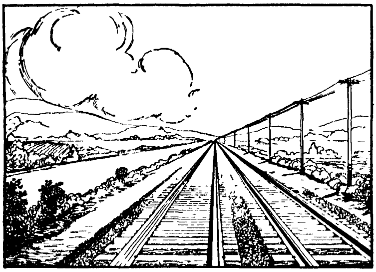


FIG. 75. LINEAR PERSPECTIVE.

are, as in Fig. 75, perceived as a series of oblique angles. Linear perspective is employed as one of the most reliable means of representing distance in pictures. In immediate perception it is a safe index of distance when the objects are familiar.

A third factor in visual localization is technically called *parallax.* The position of an object in the visual field is always perceived either as relative to the observer, or to some other object, or, again, to both. In the last case if the head is moved while there are objects in the field of view, some of which are nearer than others, the nearer objects seem to move across the field with reference to the more remote ones. The motion of the nearer object is in the direction opposite to that of the head; when the head is moved to the right the object seems to move to the left. If an object in the foreground is fixated, it is the

background which seems to move, and then the remote object seems to move in the same direction as the head (see pp. 263f.). This displacement of objects when the position of the observer is changed is an instance of parallax; it is called *monocular parallax* when only one eye is being used and thus is in contrast with *binocular parallax* which depends upon the difference in position of the two eyes with respect to each other.

The direction of the apparent motion establishes a very clear relief of foreground against background. When all other signs fail, one may, by shifting the position of the head, determine this fundamental spatial relation. The speed of the apparent movement due to parallax depends upon the distance of the objects. With a constant amount of head motion there is a rapid, and large, apparent movement of near objects and a slow, and small, movement of remote objects. It follows that the speed and amount of the perceived motion afford an indication of the *relative* distance of objects. And as the retina is extremely sensitive to those changes which signify motion (see p. 263), parallax is an extremely useful item in estimating distance.

A special case of parallax, of importance under the conditions of modern life, occurs when the whole body is moved by mechanical means, as happens in trains, automobiles or airplanes. Under these conditions the difference in rate of apparent motion of observed objects (while the rate of transportation remains nearly constant) is a safe index of their distance. Poles by the roadside flash by rapidly, trees in the middle distance slowly, while hills as much as a mile or two away may seem to stand still.

A fourth condition is *aerial perspective.* Remote objects are less distinct than near objects. Dust or smoke or mist in the air reduces the distinctness with which objects can be seen. Hazy objects are therefore perceived to be more distant than clear objects. Moreover, when an object is seen at a considerable distance, the angular size of the smaller parts becomes so

minute that these details are no longer visible. The absence of detail is then taken to be a sign of distance. If a pair of field-glasses or a telescope is used, the image is magnified and details which were not visible to the naked eye are revealed. The magnified object appears to be nearer, not only because it occupies a larger part of the retina (according to the prin-

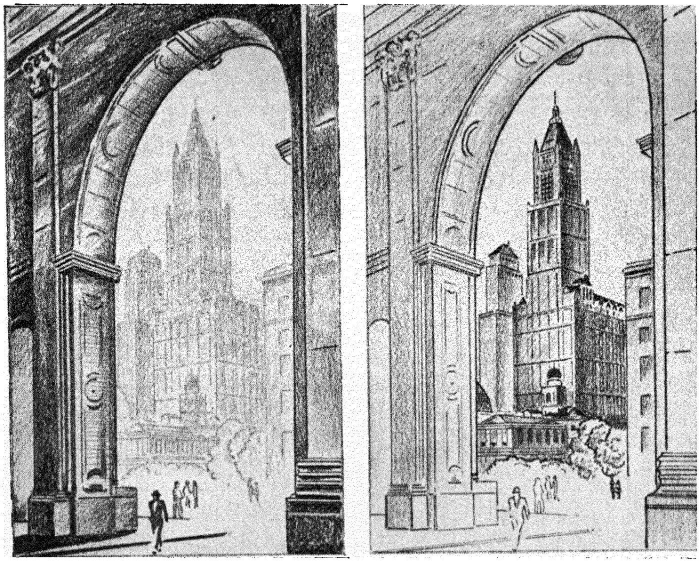


FIG. 76. AERIAL PERSPECTIVE.

ciples of linear perspective) but also because its details are more distinctly seen, as in Fig. 77.

In regions where low humidity and lack of smoke leave the atmosphere very clear, it is extremely difficult for the inexperienced individual to estimate the distance of remote objects. Hills which seem only a few miles off may prove to be many times as far away. As the very clear atmosphere reveals details of form, it frequently happens that large trees are seen as bushes, large rock formations as much smaller ones, etc., and under such conditions mistaken estimates of distance are made in terms of linear perspective. What appear to be small bushes

on the remote hill in Fig. 78 are really trees like the one in the insert.

Of more importance is the effect of *shadows*. Since light normally falls from above downward upon the objects we see,

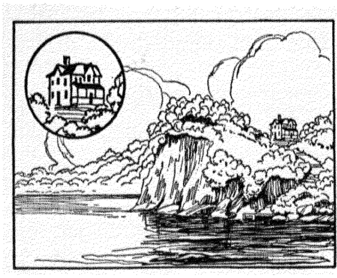


FIG. 77.

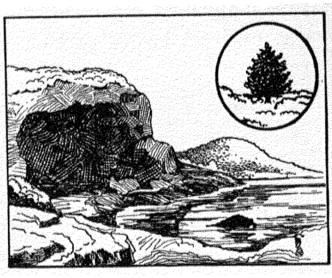


FIG. 78.

it follows that the shadows of objects are found to fall in characteristic configurations. By reversing the relations of light and shadow in such configurations, it is evident that the shadow-pattern is an important factor in determining the re-

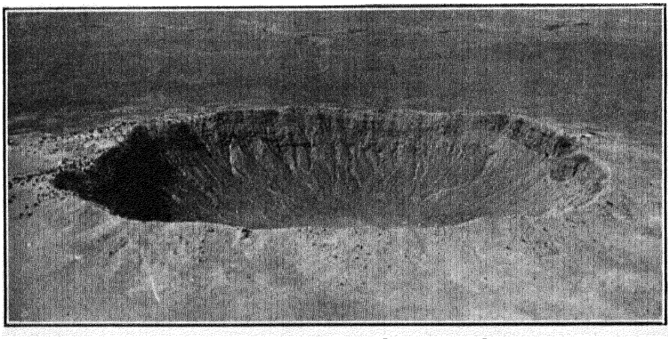


FIG. 79. THE INFLUENCE OF LIGHT AND SHADE.

The effect of reversing the relation of light and shade may be seen by turning the figure upside down. Reproduced by permission of Western Air Express Corp.

lation of foreground and background, as can be seen by turning Fig. 79 upside down. Even when the light falls from one side and not from above, the presence of shadows on the background indicates that the foreground objects (*i.e.*, those seen as complete against an interrupted field) are in relief.

There is finally a condition which is effective for either monocular or binocular vision and which arises from the fact that the eye can be adjusted so as to bring the rays of light reflected from the stimulus into focus on the retina. This adjustment of the eye is called *accommodation.* Near accommodation, it will be recalled (see p. 65), is effected by a contraction in the ciliary muscles which produces an increase in the convexity of the lens. When the ciliary muscle is relaxed, the normal eye is accommodated for any distance greater than a few yards. In this condition the eye has a universal focus analogous to that of a box camera with the lens in a fixed position relative to the sensitive plate. Within a small range

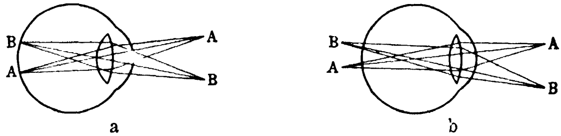


FIG. 80. THE ACCOMMODATION OF THE EYE.
In *a*, with more convex lens, the images are in focus on the retina. In *b* the focus lies behind the retina so that the images intercepted by the retina are out of focus.

of distances accommodation affords an index of distance, partly from kinesthesis, and partly from the relative distinctness of the perceived objects. The exact limits cannot be stated precisely, but for distances ranging from a few inches to not more than two or three yards the relative distance of objects is indicated if one object is in focus and another out of focus. It is evident that, when the eye is accommodated for a far object (Fig. 80, *b*), the image of any nearer object will be out of focus and blurred. By contrast with such a blurred object all the distinct objects seen with relaxed accommodation lie in the background. When it is necessary to distinguish the distance of two objects both of which lie within the range of voluntary accommodation, we must rely upon a tentative or exploratory adjustment by means of which we ascertain whether the indistinct image can be clarified by increasing or decreasing the effort of accommodation.

It is because accommodation is not the same for different wave-lengths of light that different colors, *e.g.*, the colors of figures in stained glass, may seem to be at different distances.

Binocular conditions of the perception of distance. There remain a pair of conditions which derive from the fact that the two eyes function together as one. We shall begin our

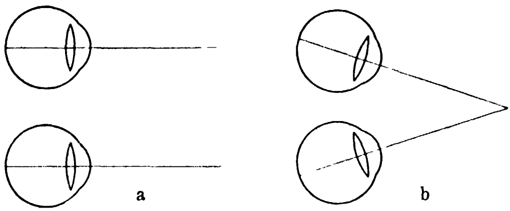


FIG. 81. LINES OF SIGHT

In *a* of the figure, the lines of sight are directed to the horizon. In *b*, they are converged upon a near object

treatment of them by considering them purely physiologically. The first is the mechanism of convergence.

Normally the muscular systems of the eyes are so adjusted that when an object is observed at the horizon the two lines of sight are parallel (Fig. 81, *a*). But when the object is near, the muscles effect a movement of convergence by means of which the lines of sight converge upon the object (Fig. 81, *b*). In each eye, when convergence has been accomplished, the image of the object which is being observed falls upon the region of clearest vision (the fovea). If the object is of large extent or if there are several objects being viewed, the projection will extend beyond the fovea. When such objects are seen entirely as single, their images are said to fall upon *corresponding* or *identical points* in the two retinas, like the images of *A* and *F* in Fig. 82. Corresponding points are not

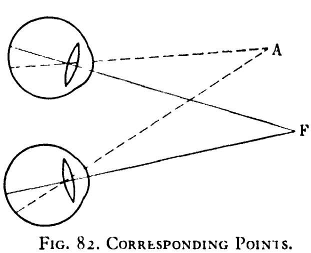


FIG. 82. CORRESPONDING POINTS.

anatomically identical, for if the image of *F* falls on the fovea in each eye, the image of such a point as *A* will fall to the nasal side of the left and to the temporal side of the right retina. The images of objects farther or nearer than the point of fixation upon which the eyes are converged cannot fall on corresponding points. The points upon which they do fall are called *non-corresponding*; and if we attend to such objects, without changing the point of fixation, they are seen as double. This, or any other condition resulting in seeing two objects where there is only one, is called *diplopia.* The images of points nearer (Fig. 83, *a*) or farther (Fig. 83, *b*) than the point of fixation

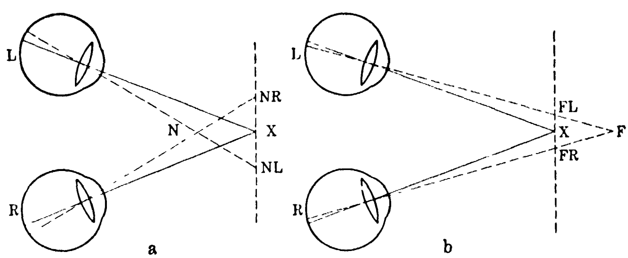


FIG. 83. DOUBLE IMAGES.
In *a* the images of *N*, a point nearer than the point of fixation, *X*, are seen at *NR* and *NL*. In *b* the images of *F*, a point farther than *X*, are seen at *FR* and *FL*.

are called *double images.* When they belong to a *nearer* point they are said to be *crossed* because the image in the *right* eye (observed when the left eye is closed) lies to the *left* of *X* (Fig. 83, *a*), and *vice versa.* When the double images represent an object farther than *X* (Fig. 83, *b*) they are said to be *un-crossed, i. e.,* the *right* eye's image is on the *right* of *X* and the left eye's image is on the left.

The facts of convergence help us to understand why it is that having two eyes we nevertheless see fixated objects as if we had a single eye—the two eyes functioning as one. But how far convergence alone can aid us in the perception of the distances of objects is a question. The movements of the muscles controlling the eyeballs in convergence are normally bound up with the movements of the ciliary muscles in accommoda-

tion. If a card is held in front of one eye while the other is accommodated for a near object, it will be found, on removal of the card, that the other eye is properly converged upon the object. Thus convergence and accommodation are inseparable in normal vision, and the importance of the movements of convergence as a separate factor in depth perception is not easily determinable. In any case, experiments seem to show that convergence cannot be effective for depth perception at a distance greater than 20 meters.

The other condition for the visual perception of distance, which arises from the use of the two eyes as one, is called *stereoscopic vision*. Since the two eyes are not in exactly the

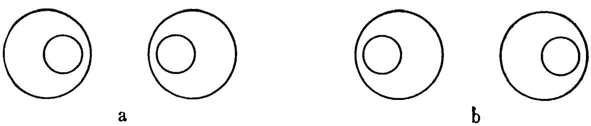


FIG. 84. A TRUNCATED CONE AS SEEN BY EACH EYE SEPARATELY.
The small circles represent the small ends of the cones, and the large circles represent the large ends. In *a*, the small ends are nearer the observer, and in *b* the large ends are nearer.

same place, it is evident that they obtain somewhat different views of nearby objects. If a small truncated cone made of paper is placed with its base against the wall directly in front of the observer, the two eyes will obtain respectively the views shown in Fig. 84, *a*. In the figure the base and plane of truncation are represented by circles, and the surface (exterior) of the cone is not shown. If the cone is attached to the wall by its small, truncated end, the respective views for each eye are as represented in Fig. 84, *b*. Here the circles still represent the ends; but the surface, which is not shown, is now the *interior* surface. When the actual cone is looked at with both eyes we fail to note that there are two views but we obtain a unique perception of the relative distance of the base and apex. In the arrangement of Fig. 84, *a* we see the small end jutting out toward us. In the arrangement of Fig. 84, *b* it is the base which juts out. The unique effect of solidity or tridimensionality

which is thus obtained is one instance of stereoscopic vision, *i.e.,* the seeing of solid forms.

The perception of objects as solid means that the object has *depth*—projects in the third dimension. The perception of solids, therefore, involves the perception of distance. It is apparent that this perception is in some way connected with the difference in the appearance of an object as seen with the two eyes (binocular parallax). When the paper cone of Fig. 84 is observed with the left eye (while the other is closed) the view of the *exterior surface* is what is represented in the left view of Fig. 85, *a.* Since the left eye is actually located somewhat on the *left side* of the object, the view obtained by that eye will

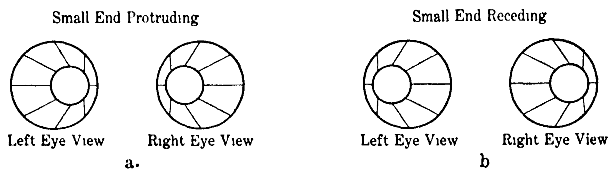
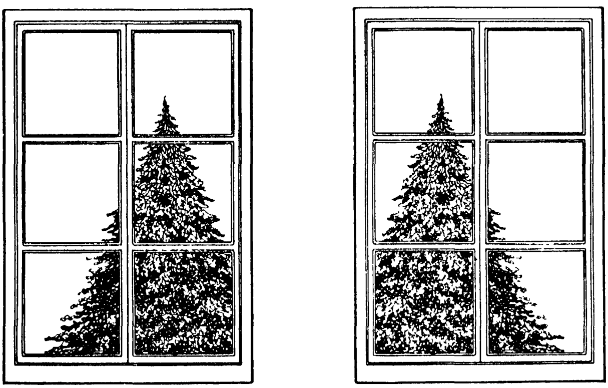


FIG. 85. THE TRUNCATED CONE OF FIG. 84 WITH ITS SURFACES REPRESENTED BY STRAIGHT LINES.
The figure is described in the text.

show more of the *left surface* of the object, as contrasted with the view seen by the right eye in which the right side predominates. When, on the contrary, the cone is placed with its small end against the wall, it is the *interior surface* which is visible, and in this position (Fig. 85, *b*) the left eye sees more of the *right side* than does the right eye. The facts are more clearly evident if one sits a short distance from a window casing and considers some object such as a tree outside the window. When the left eye is used (Fig. 86, *a*) more of the right side of the remote object is seen than when the right eye is used (Fig. 86, *b*). By alternately closing first one eye and then the other, while the remote object is fixated, the nearer one (in this case the window frame) will be seen to jump back and forth from left to right. The difference in position of the eyes gives *parallax*; but the difference between the two views is due to *retinal disparity.*

Two objects can be distinguished as separate when their angular separation is remarkably small (of the order of magnitude of 1 minute of arc or less). Double images afford a reliable criterion of distance when their displacements from the single image of the fixated object are of a similar magnitude, so that, although the distance between the eyes is small, binocular parallax is obtainable when the object is 1000 yards or more distant. Moreover, retinal disparity is present whenever some one object such as a window frame, or a part of an



a b

FIG. 86 BINOCULAR PARALLAX.

automobile or of an airplane, is within the field of view, so that stereoscopic vision is almost always present in normal vision.

Retinal disparity means that the retinal images are different (like the drawings in Fig. 84) and that geometrically regarded they fall not on corresponding but on non-corresponding points. We might therefore expect to see double images. Instead we see a single object in three dimensions. Just how the organism accomplishes this feat we do not know, but there is a synthetic experiment which proves it. We may take two drawings of a solid object as seen by each of the two eyes (Fig. 84) and separated by the interocular distance, or two photographs made by two cameras placed side by side in a position corresponding to that of the two eyes and separated by the same

distance. We may then view either the plane drawings or the photographs in a stereoscope. The stereoscope is an instrument which makes it possible to combine the two pictures by presenting to each eye the picture that would be seen by that eye. When the observation is made we see not two plane pictures, or a single plane picture, but a tridimensional solid object in plastic relief. The plastic effect can be enhanced by increasing the distance between the two cameras, and in the production of commercial stereoscopic pictures this is the usual practice. Increasing the distance between the cameras also reveals details of relief (as in topographical pictures for map-making) which are too slight to be perceived by the natural eye.

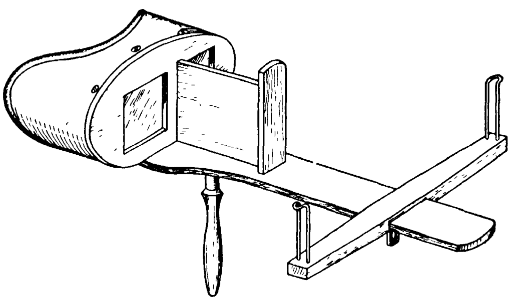


FIG. 87. BREWSTER STEREOSCOPE.

The stereoscope itself as an instrument has only two essential features: (*a*) there must be a partition so that neither eye will see the picture intended for the other eye, and (*b*) there must be some device which will move the image of each picture to the same place on each retina that would be occupied by the images of the real object if the two eyes were seeing it. The last-named result is obtained in the common Brewster stereoscope by means of prisms which project the images on to the retinal areas upon which they would fall if the eyes were looking along parallel lines, as at distant objects, while actually the eyes remain converged upon the cards in the instru-

ment. The Wheatstone stereoscope, an older form, employs mirrors instead of prisms.

Visual localization. In visual perception we localize objects as in front, above or below, to the right or left hand. Every one of these places or directions has its own special conditions. When an object is localized as directly in front of the observer the body is usually upright, the eyes are directed forward and converged upon the object and the image of the object falls upon the fovea. When we sight an object by bringing it into line with another object, we may use either a single eye or both eyes, and we move the body or head until the two objects are in line with each other and the eye or eyes. Some individuals sight with only one eye although both eyes may be open, and the eye thus used is called the dominant eye. Other individuals sight with both eyes, the line of sight passing between the two eyes. When it exactly bisects the interocular distance, both eyes are said to have identical visual direction, *i.e.*, the line of direction is identical in function with the line of sight of an imaginary single eye called the Cyclopean eye. You may test yourself in this respect by sighting with both eyes open a distant object and a pencil held at arm's length. When the two objects are in line, first close one eye and then the other, noting in both cases whether the pencil is still in line with the distant object. If this occurs with either eye, that eye is dominant. If, however, the pencil lies to the left of the distant object when the left eye is closed, and to the right when the right eye is closed, neither eye is dominant and the line of sight passes between the two eyes.

When we perceive an object as in front and its image falls on the fovea, the image of objects above the fixation point will fall below the fovea, and conversely the images of objects below the fixation point will fall above the fovea. Similarly, images of objects on the right side of the fixation point will fall on the left, and images of objects on the left of the fixation point will fall on the right of the fovea. In every case the visual angles subtended by any pair of stimuli and by their

corresponding retinal images are the same. This fixed relation between stimuli and their retinal images is the first condition for the perception of the position of objects in the visual field.

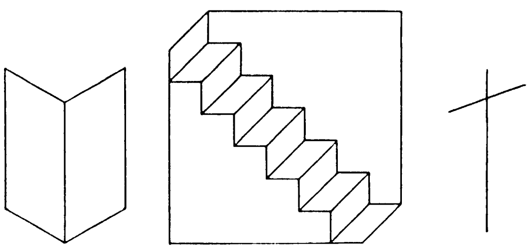
The second important condition is the position of the body, the head and the eyes. We may perceive an object as above by raising the eyes or by tilting the head backward. Conversely, we may perceive an object as below by lowering the eyes or by tilting the head forward. Similarly we perceive objects as to the right or to the left by turning the head or the eyes (or both) to the right or to the left hand as the case may be. When the head and eyes are turned to the right or to the left, localizations in front are often referred to the front side of the body.

These two sets of conditions may and do reinforce each other, but at times they conflict. When, for example, we lie flat on our backs the ceiling or sky is by eye in front, but by position of the body it is above. When we lie on a side of the body, directions are still more confused. It is said that aviators when emerging from a cloud sometimes discover that they are flying upside down by perceiving the sky below and the ground above them. Under the exceptional conditions of flying, the somesthetic cues for the perception of the posture of the body were not effective, and the cloud in which they were flying provided no visual cues.

The visual perception of direction is particularly important for the guidance of the legs, arms, hands and fingers in locomotion, in steering boats and automobiles, in reaching and grasping, in writing and drawing and in many other bodily movements. The direction of movements forward, up and down, right and left normally corresponds to the visual directions front, above and below, to the right or left hand. When, as in steering a rowboat with a rudder, the handle of the rudder must be pushed in a direction opposite to that in which the boat is desired to turn, the movement at first seems unnatural but is soon learned. Also in dressing before a mirror, we have long since learned to adapt our movements automatically to the reversed directions of the mirror-image; we generally even fail to notice the changes in direction until we rely

upon the mirror-image as guide for unusual movements. In a famous experiment lenses which re-inverted the retinal images, and made objects look upside down, were worn continuously during waking hours for several days. It was found that, whereas the experimenter at first made movements in the wrong directions, he quickly adapted them to the new visual appearances.

Illusions of size and direction in visual space perception. Visual perceptions of size, direction and distance do not always accord with the knowledge of them as it is gained



a b c

FIG. 88. ILLUSIONS OF REVERSIBLE PERSPECTIVE.

from other sources. Such perceptions are referred to as *illusions*. In the great majority of cases illusions arise in highly artificial situations and can be traced to the incomplete, and therefore unnatural, integration of the data to which the organism is required to react.

Illusions of 'reversible perspective' occur when some or all of the criteria of relative distance are lacking or, if they are present, are in contradictory relations. Thus in line drawings, such as those of Fig. 88, where there is no linear perspective or other criteria of depth, the relation of foreground and background is undetermined. Such drawings are *ambiguous* as to perspective.

In the sky the beam of a searchlight will appear to come to a point and its edges will seem to converge even though they are really parallel, like the lines in Fig. 75. The illusory appear-

ance of parallel edges can be created by spreading the beam in a slightly divergent fanlike pattern. But the amount of the divergence depends upon the apparent distance of the part of the sky upon which the beam falls and, as this is indeterminate, the effect is inherently illusory. The perceptions cannot possibly correspond to the 'facts.'

Failure to represent, in drawing, the surface of a solid object, leaving parts of the background visible, which, in a natural object, would be obscured by objects in the foreground, likewise gives ambiguous perspective (Fig. 89).

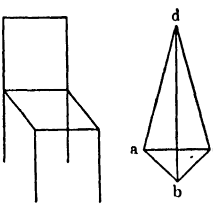


FIG. 89. AMBIGUOUS PERSPECTIVES.

When angles are introduced into drawings, without a supporting context of other cues as to direction or distance, illusions also occur (Fig. 90).

A considerable group of illusions arises from our inability to abstract a part of a linear figure from the integrated whole in which it is embedded. The line *AX* in Fig. 91, *a* appears longer than *AY* because it is the diagonal of a larger parallelogram. The apparent size of the circle in Fig. 91, *b* depends upon the size of the entire figure. Similarly, line *c* of Fig. 92, *a* appears longer than *d* because it is part of a larger total figure.

The difficulty of isolating linear distance from a total figure becomes even more apparent when the distance itself is not represented by a line. In Fig. 92, *b* the distance which *separates* *B* and *C* is equal to the distance which *includes* *A* and *B*, but it cannot be made to appear so. This principle is operative even when it entails contradictions within a larger total figure, as in Fig. 92, *c*.

On the other hand, *contrast* between adjacent elements of figures may result in an illusion of the size of the other elements in the figures to which the compared parts belong, as in Fig. 93. Here the part (and consequently the area) is apparently increased when one part is magnified by contrast with an adjacent small part.

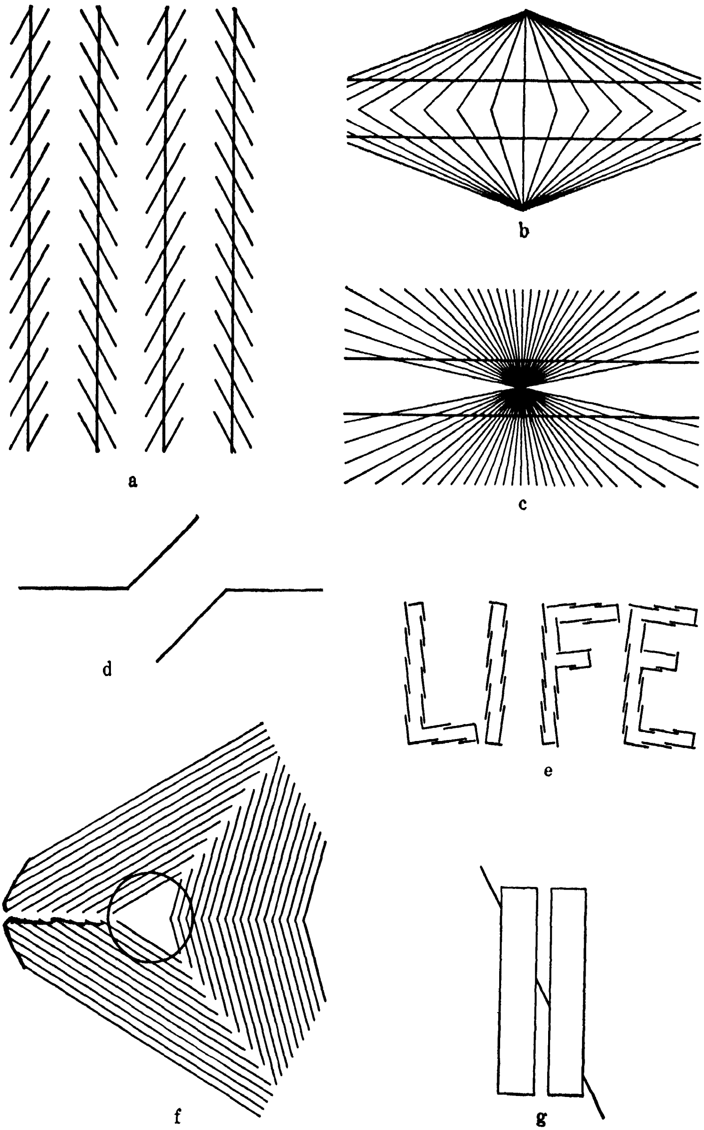


Fig. 90. Angle Illusions.

Although it is true that most of the more striking illusions of visual space perception are found in artificial situations in which the conditions of normal perception are distorted or partly lacking, nevertheless many instances of illusions arise in wholly natural situations. *Irradiation* occurs when an area is strongly illuminated, and as a result a surface or object appears larger when it is bright than when it is dark. It is generally assumed that the effect of the relatively intense stimulus actually spreads in the retina beyond the geometrical boundaries

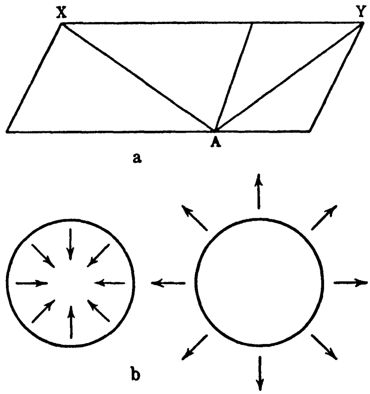


FIG. 91. INFLUENCE OF A LINEAR FIGURE TAKEN AS A WHOLE UPON THE SIZE OF ONE OF ITS PARTS.

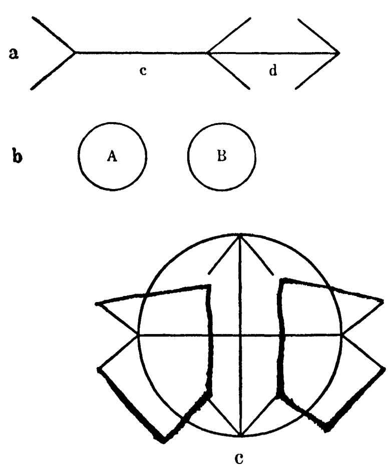


FIG. 92. EFFECT OF A TOTAL FIGURE UPON AN ISOLATED LINEAR DISTANCE.

of the image. Another illusion which occurs regularly under ordinary conditions arises whenever vertical distances are com-

pared with horizontal ones. A vertical distance appears to exceed a horizontal distance which is really the same. A somewhat analogous illusion appears when one looks vertically downward from a height; the vertical distance is overestimated, and the angular (horizontal) magnitude of objects on the ground is correspondingly underestimated.

An instance lying between the artificial and the natural types of illusion is presented by the fact that a divided or filled area appears to be more extensive than an equal area which is not filled in. In Fig. 94, *AB* appears to be longer than *BC*.

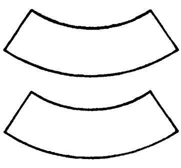


FIG. 93. SPATIAL CONTRAST.

Theory of visual space perception. Although the various spatial relations of objects have been considered separately it should be realized that an object is at once perceived with a certain degree of distinctness, at a distance, of a size, and in a certain direction. Furthermore the conditions of these spatial relations are interrelated so that no one of the relations can be considered without involving the conditions of the others. For example, we have seen that the perception of the size of an object is in part a function of the perception of its distance and of its details.

Some of the conditions are fundamentally optical; they are based upon the eye regarded as an optical instrument. Others are muscular; the eye is itself a moving instrument and it also moves as the body or head moves. Still others are neural; the receptors in the retina, the optic nerve and its connections in the midbrain and in the visual cortex are all involved. These three sets of physiological conditions constitute what we may call the optical system. Throughout our discussion we have regarded this system as functioning normally, but a derangement anywhere in it affects perception.

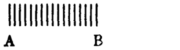


FIG. 94. FILLED AND EMPTY SPACE.

Although we cannot ignore the part that the nature of the stimulus plays, a theory of visual spatial relations must ultimately be stated in terms of the optical system. Our knowl-

edge of the eye as an optical instrument and of the action of the muscles which move it is perhaps sufficient to permit the statement of an adequate theory; but we know so little about the neural part of the system, particularly about the function of the cortex, that an adequate theory is at the present time impossible.

There are, however, many facts which point directly to the function of the neural part of the system. The receptors of the retina are the rods and cones. These are arranged in a mosaic unevenly distributed over the retina. We might expect, therefore, that their stimulation by a surface such as a sheet of white paper would result in a perception of spots of white with nothing between them, but instead we perceive a continuous surface. This spatial continuity may result from a spread of excitation at the retina, or from the function of the optic nerve, or it may be of the cortex.

Again, there is the large number of facts that are generally credited to 'experience.' The dependence of the perception of relative distance upon the known size of objects, upon linear perspective, upon atmospheric conditions; the dependence of the perception of relief and intaglio upon the distribution of light and shade; the dependence of seen directions upon bodily posture; the part that attitude plays in the perceptions of size and distance—all these and many more show the effects of past experience. But these effects are something more than the mere accumulation of knowledge from which we consciously or unconsciously infer the size, distance or direction of objects. The perceptions are immediate; generally they are compelling in the sense that we cannot see them in any other way, and it does not in the least change the perception to know its conditions. This factor of experience is one common to perception at large and will be discussed later (see p. 294).

Finally, there is the fact that conditions may reinforce each other or conflict with each other. They reinforce each other, for example, when binocular disparity, both linear and aerial perspective, interposition and known size of objects unite in conditioning the perception of relative distance. They seem

to reinforce each other in such a way as to make the perception more certain, more compelling and more stable. They conflict with each other when, for example, in 'size constancy' the known size conflicts with the law of the visual angle, or when bodily position conflicts with retinal image in the perception of direction. The results are variable; but they are revealed whenever the perception is uncertain, ambiguous and instable.

LOCALIZATION OF SOUND

One of the primary functions of the ear is to give warning of approaching dangers. This function is served to some degree even when the sound is not localized. Loud sounds are disturbing to infants and produce in adults a characteristic startle reaction. If, however, the sound can be localized as to distance and direction it becomes possible for the individual to orient himself so that the eyes can be brought to bear on the locus of the sound.

Auditory perception of distance. The distance away of the source of a sound is perceived principally by reference to its relative loudness. A near sound is loud; a distant sound is faint. Loudness in this context is, however, probably something more than mere intensity. The distant sound is not only fainter but also of less volume and, especially for noises, of a different timbre. By reference to these differences we learn how the sound of a particular object varies with different distances, and a particular sound is eventually perceived as far or near. The perception may, of course, be made more uncertain by the presence of a background of accompanying sounds. This background varies greatly in intensity, and the apparent intensity of a particular sound is affected by the masking effect of the background.

The direction of sounds. In the laboratory experiments upon purely auditory localization the results of which have been set forth in an earlier chapter (see pp. 129-133), all stimuli except sounds were of course eliminated. In everyday life there

are frequently other cues to localization. Chief among these are the visual identification and localization of the object that produces the sound. If the sounds are familiar the visual localization of the object that we regard as the source of the sound determines the localization of the sound itself. In an experiment, ear trumpets were arranged on the head in such a way that each ear was connected with a trumpet on the opposite side of the head, thus in effect putting the right ear on the left and the left ear on the right side of the head. Under these conditions a voice coming from a speaker on the right was first auditorily localized on the left; but when the speaker was *seen*, the localization of the sound changed to the right.

Similarly under natural conditions visual localization is a constant aid to sound localization. By ear alone, it will be recalled, we cannot localize a sound above or below the horizontal plane in which the head lies. Nevertheless when we can see the object its seen place determines the place of its sound, whether it be that of an airplane above us or of a dog at our feet.

In laboratory experiments localization is usually determined verbally, or by pointing, while the head remains in a fixed position. Such an arrangement is highly artificial. Normally, localization is an incidental result of re-orientation of the body (or at least of the head) as a result of which time, phase and intensitive differences between the ears are reduced toward zero by turning the face toward the source of sound. Such a movement of the head may be governed by simple reflexes and may even be thought of in terms of the tropistic responses which are found in insects and in very lowly organisms. A tropism results in a re-orientation of the organism such that the effects of stimulation on the lateral surfaces are equalized. The body is turned so that its long axis is in the line of the light, sound or electrical force. If the tropism is 'positive' the animal brings its anterior end toward the source of energy; it faces the stimulus. The behavior of a free animal in facing in the direction of a source of sound, thus equalizing the effects

in the two ears, is at least strongly reminiscent of primitive tropisms.

LOCALIZATION OF SMELL, TASTE AND ORGANIC SENSATIONS

In man the olfactory organs are so situated that stimulating substances must pass through tortuous channels before reaching the special sensory cells, and the functioning of these organs alone is not effective for localization. However, the source of an odor can often be detected with the help of exploratory movements because intensity increases on approach; this permits some degree of bodily orientation toward the point from which the odor emanates. For animals such as dogs, and for some ants and other organisms, odorous traces left upon the ground constitute important cues for the guidance of movement.

Taste is almost exclusively interoceptive in function and does not afford any basis for perception of exterior space. Localization, in so far as it occurs, is on the surface of the tongue. No systematic experiments in taste localization have been made. When, however, drops of solution impinge upon a papilla, the taste quality excited is localized at the place of contact. It seems that with equal pressure over the entire surface of the tongue, as when the mouth is filled with a liquid, taste qualities can still be referred to specific areas of the tongue.

In general, the interoceptors are of little or no importance in the discussion of space perception. Organic sensations are localizable within the spatial framework of the body as it is visually and tactually known. At no time are we aware of our bodies as a pattern of organic sensations alone, although they may enter into a pattern which is predominantly tactual and kinesthetic.

Pains arising from rheumatic joints and sore muscles are more or less correctly referred to their points of origin. Distension of the walls of the intestine by gas gives rise to a

quality which is generally referred to the interior of the abdominal cavity. The pains of appendicitis are generally referred correctly to the right side of the trunk. Heart pains are referred to the general region of the heart. Localization of such pains is often sufficiently accurate to enable the diagnostician to discover the locus of the physiological disturbance. On the other hand, pains are frequently experienced in a part of the body which is physiologically normal and in some cases it can be shown that they are caused by pathological conditions in some organ remote from the apparent position of the pain. A pain in the shoulder, for example, may be caused by a condition in the heart, one in the leg by a condition of the bladder. Such 'referred pains' may be of value to the diagnostician because there is a definite connection between the apparent seat of the pain and the locus of its cause, but to the patient himself the localization, although incorrect, is as real as any other localization.

BODILY MOVEMENT AND POSTURE

The perception of space implies much more than a mere passive apprehension of the relative positions and distances of objects. In perceiving spatial relations the organism is concerned with moving about or getting ready to move about among the objects which it perceives. Localization of a point on the skin is significant because the hand can then be moved to that point and the involved part protected or the stimulus removed. Sounds or odors are localized to enable the organism as a whole to move with respect to them. In visual space perception, movements of the head and of the eyes themselves are directly involved and their final significance must be stated with respect to those movements which are to be made in approaching objects or in bringing them into a position where they can be better seen or handled.

The execution of movements of any kind implies, however, some cognizance of the position from which the movement starts and some means of gauging its force and extent. This

is the function of the proprioceptive mechanisms of kinesthetic sensation in the joints, muscles and tendons, and in the organs of the semicircular canals, utricle and saccule of the inner ear (see pp. 173-184). The mechanisms of the joints, tendons and muscles enable us to distinguish amounts of weight, pull and resistance. The thresholds for the discrimination of weight with active movement of the limb are much smaller than the thresholds for passive pressure. These mechanisms enable us to estimate fairly accurately the extent of voluntary and of passive movements of the limbs. The thresholds for angular displacement at the joints are of the general order of only one degree of arc or less. But even more delicate adjustments are involved, for example, in holding a rifle so as to hit a target a yard wide at a distance of a thousand yards. In such situations, of course, visual perceptions supplement the proprioceptive cues in guiding the movements and in giving notice of the direction of error.

It is expedient to refer all movements to some one line of reference, and the vertical line through the body in its upright position serves this purpose. In this position the body is in equilibrium with respect to gravitational pull. This position is maintained partly in terms of proprioceptive impulses from the feet and legs supporting the body, as is evidenced when these impulses are cut off in the disease *tabes dorsalis* which results from a lesion in the afferent tracts of the spinal cord. The vertical position is maintained partly in terms of the organs of the inner ear as is evident when these are injured or artificially stimulated. Furthermore, it can be maintained to a large extent through visual perception; tabetic patients can maintain equilibrium if they can see the ground and their limbs. For the normal individual in the dark or with the eyes covered there is a considerable increase in the involuntary movements of the trunk as the body sways and is returned to its vertical position.

The proprioceptive mechanisms are concerned with *posture* as well as with movement. The actual movements concerned in acts such as typewriting vary according to the general pat-

tern of muscular and skeletal adjustments involved in the particular posture from which the movement is executed. Such an action as rising from a seat involves different movements on various occasions, depending upon the postural dispositions of the body while seated. The other senses are also closely dependent upon proprioceptive continuity of posture. Objects in the visual field do not rock when the head is inclined to one side, but regularly remain upright. The upright position of such objects cannot be determined within the visual field itself. Their vertical axes of reference are functions of the position of the body in the field of gravity as determined by the proprioceptors. When the vertical position is artificially disturbed, vertigo results and visual objects do actually seem to rock or swim.

ORIENTATION

The ability of the organism to find its way about in the world is a commonplace of everyday observation, and yet it presents phenomena which are very obscure and for the explanation of which psychological and biological science in general have found no adequate hypotheses. A particular species of bird lays its eggs in the nests of other species on the northern plains of Canada. The parent generation leaves while the young are still in the nests of their foster-parents. Yet these young birds find their way unguided to the winter feeding grounds of their own species in South America. In some species the route followed on the return journey is totally different from that of the outward journey. That followed the second year is, for some species, different from that of the first year. It is now well established that homing pigeons make use of visually perceptible landmarks with which they have become familiar. Bees and wasps also use such landmarks. Ants sometimes employ visual cues and sometimes depend on odors. Rats in a laboratory maze are able to learn to follow a path when seemingly deprived of all exteroceptive functions. Both rats and human beings show the ability, in a

maze, to negotiate short-cuts and detours, which involve making new turns and traversing paths of different lengths, in order to reach a goal the location of which has previously been learned. Even the limpit, a mollusk, is capable of making forays and returning to the precise niche in the rock into which its shell has grown to fit.

The evidence is abundant that animals are capable of maintaining a general direction of movement with respect to nests, feeding places and other fixed positions in their environment. The actual movements involved may be greatly interfered with (by interposition of obstacles, injury to motor system, change of mode of locomotion from wading to swimming) and the animal will, nevertheless, succeed in correctly maintaining the general direction of his movements. Animals are generally correctly *oriented* in their environments. There is every reason to believe that men likewise are generally correctly oriented or possess, as it is often inaccurately stated, 'a sense of direction.' We live in a world which is, for the most part, full of guiding lines along which we move without much heed of the directions we are taking. And in a wilderness where there are no such lines we use our wits rather than our senses and guide ourselves by deliberate interpretation of consciously observed and analyzed relations.

THEORY OF ORIENTATION AND OF SPACE PERCEPTION IN GENERAL

A comprehensive theory of space perception must take account of three groups of phenomena: (*a*) the exteroceptive functions by means of which the organism is directed toward, and learns to move with respect to, sources of external stimulation, tactual, auditory, visual, etc., (*b*) the proprioceptive functions by means of which the organism maintains its posture and position with reference to gravity and the immediately supporting ground, and (*c*) the substitutional processes by means of which the organism is enabled to move correctly in the network of objects constituting its environment when

the stimuli, which have previously supported action, are no longer present.

(*a*) The traditional discussion of space perception is almost entirely confined to the phenomena of the first group, the exteroceptors, and particularly to vision. For the normal individual the conscious experiences in which space is involved are preponderantly visual. Furthermore, the intricate mechanisms of visual perception are intrinsically fascinating. And actually vision is responsible for niceties of discrimination as to distance, form and direction which cannot be attained otherwise. Vision, too, more than any other sense, presents, at a glance, in one moment of experience, a large part of the whole framework of the arrangement of the objects surrounding the individual in space.

(*b*) The body of factual information available concerning the second group of phenomena, those connected with the proprioceptors, is very small in contrast with the theoretical importance of these mechanisms. This paucity comes about partly because proprioception is not easily accessible to introspective description, and partly because the internal situation of the organs and their diffusion throughout the body make it extremely difficult to devise satisfactory experiments. Except for some experiments upon the mechanisms of the inner ear and the reflexly associated eye movements, we are compelled to rely upon clinical data, or to resort to indirect experiments, in which the proprioceptive functions are assumed to be the only ones remaining after the removal of other sensations. The latter method involves either the operative destruction of the sense mechanisms or the experimental control of their operation. Operative destruction is not feasible in man, and the results of this method with lower animals are questionable. Often the extirpation is not demonstrably complete (except for vision, where the removal of the optical organ is simple).

The meager results which have been achieved by these various methods lead to the hypothesis that the proprioceptors are responsible for at least three of the basic factors in space perception. They yield, first, the awareness of posture with respect

to gravity, and awareness of the position of the members of the body with respect to the posture of the whole body. Without this basis the notion of *direction* loses its meaning. Secondly, these organs make possible control of the direction of movement, right, left, up, down, etc. And, thirdly, there must be some temporal integration of movement, so that the animal is able to discriminate distances traveled. It is frequently asserted that this integration involves both distance and direction, so that, at the end of a series of movements, not all in the same direction, the animal is aware of the net distance traveled in the straight line connecting the starting point with the point of arrival. Although the facts of behavior often seem to demand some such compound integration of direction with distance, it is difficult to conceive of any process by which it could be effected.

(*c*) Finally a theory of space perception must account for the cooperation of the proprioceptors and exteroceptors so that the organism can control its movements, not merely egocentrically, but also with respect to the positions of things seen, heard, felt, smelled, etc. In this outer world the organism is itself one of the objects existent in space. The feel of objects on the skin corresponds with the movements used in touching those parts of the skin; the direction of objects seen corresponds with the movements, to the right or left or up or down, used in reaching those objects; the distance seen or the loudness of a sound corresponds with the amount of movement needed to bring the body of the organism to the spot. The various sensory 'cues' from sight and hearing and touch and smell are similarly interrelated among themselves. The *same* empirical space can be perceived in all these ways.

REFERENCES

1. Bourdon, B. *La perception visuelle de l'espace* Paris: Schleicher, 1902.

2. Fröbes, J. *Lehrbuch der experimentellen psychologie.* Freiburg: Herder, 1923. I, pp. 257-378.

3. Guillery, H. Sehscharfe. In *Handbuch der normalen und patho-*

logischen Physiologie. Berlin: Springer, 1931. XII (part 2), pp. 745-811.

4. HENRI, V. *Ueber die Raumwahrnehmungen des Tastsinnes* Berlin. Reuther and Reichard, 1898.
5. HOFMANN, F. B. *Die Lehre vom Raumsinn des Auges.* Berlin. Springer, 1920-1925.
6. KATZ, D. *Der Aufbau der Tastwelt.* Leipzig: Barth, 1925
7. LADD, G. T., and WOODWORTH, R. S. *Elements of physiological psychology.* New York· Charles Scribner's Sons, 1911.
8. RIVERS, W. H. R. Vision. In E. A. Schafer's *Text-book of physiology* Edinburgh: Pentland, 1900. II, pp. 1122-1148.
9. TSCHERMAK, A. Augenbewegungen. In *Handbuch der normalen und pathologischen Physiologie* Berlin: Springer, 1931. XII (part 2), pp. 1001-1094.
10. TSCHERMAK, A Optischer Raumsinn. *Op cit* (9) XII (part 2), pp. 834-1000.

Chapter 10

TEMPORAL PERCEPTION

Temporal perceptions are integrations of a number of events which occur successively. Thus they are always analogous to spatial perceptions in which simultaneous occurrences are integrated into a spatial pattern. All time perceptions have two characteristics in common.

In the first place they are in their very nature continuous; events 'durate,' they go on. They have a beginning and an end, but once begun they last until the end is reached. This durative character is a common property of all sensory experience. It has been called protensity or the attribute of duration, and is the descriptive characteristic of mental time.

The temporal perceptions also have an important functional aspect, that is to say, successive events are said to be integrated because they act together as parts of some single whole—in memory or in constituting a situation which is the occasion for action or thought. A melody is an integration in time: the first few notes of a measure mean the whole measure and almost inevitably lead to its completion in thought. The significance of the separate notes depends upon the context of the preceding and succeeding notes. A joke is a similar temporal integration, for it has a definite structure that happens in time; there must be in the joke a development of thought in a given direction, a sudden dislocation of the trend and then a resolution of the movement of thought in a new direction. The joke does not really have a 'point,' but a pattern that runs itself out in a definite time.

Because a joke or a melodic phrase has no meaning at any single point of time but only in its temporal entirety, it has

This chapter was written by Miles A. Tinker of the University of Minnesota.

sometimes been supposed that time psychologically telescopes itself, so that the immediate past is somehow carried on into the immediate present. Logically the conception of the present is specious, because from this point of view the present is nothing more than the critical point which separates the future from the past; it has no appreciable duration in its own right. Nevertheless, consciously to all of us the present exists: experience, as we know it, seems to lie entirely in the present, although logically the present is of infinitesimal duration. How is this paradox explained?

The explanation lies in the fact that both science and common sense recognize the existence of temporal integrations. The conscious present is the length of time that an integrated whole—a musical phrase or a joke—takes; it is the duration of any series of events that depends upon its temporal completion for its proper psychological functioning. The past persists into the future only by way of its effects, but the fact that we all know that temporal integrations exist has led to our common belief in a present of variable but finite length.

These two characteristics, the durative nature of experience and the integration of a number of successive events into a single whole, are the basic conditions of four principal types of temporal perception: the perception of continuity, the perception of succession, the perception of temporal length and the perception of rhythm.

THE PERCEPTION OF CONTINUITY

The perception of continuity occurs under two different circumstances. In the one an experience is running its course without appreciable qualitative or intensitive change; in the other the experience is itself perceived as changing. With the former we tend to ignore the durative character of the experience until some change occurs in the total situation; until, for example, a steadily burning lamp begins to flicker, until the pressure of our clothing gets uncomfortable, until a sound is lasting longer than we expected, until we fear a particularly

pleasant experience may cease, until we begin to wish a toothache would stop, until we wonder, as we listen to Niagara, if its roar will ever be stilled. In all such cases we perceive, what otherwise escapes our notice, that the experience of the moment is continuously going on even though no change in the experience is observable.

The perception of continuity is much more obvious when the experience of the moment is itself changing; when, for example, the colors of a sunset slowly fade, when a whistle rises or falls in pitch, or grows louder or softer in intensity, when the strain of effort increases or diminishes, when with adaptation a hue loses saturation or a pressure gradually dies out. In all these instances we perceive a continuous change in quality or intensity. We may also perceive a similar change in the position or in the size of an object as when a bird flies across the field of vision, or when we see an object growing larger or shrinking in size.

PERCEPTION OF SUCCESSION

In the perception of succession, stimuli are presented at a rapid rate and qualities or events are perceived as discrete and as following one another in time. The rate of presentation, however, must not be too rapid, for then the qualities will fuse into a single continuous impression. If, for example, black and white sectors on a color wheel succeed each other too rapidly in rotation, one perceives either a smooth gray or else, if the rate is a bit slower, a continuous flicker. With auditory stimuli and a rate that is too rapid, one perceives either a continuous noise, or else, if the rate is slower, a harsh, rough noise. These phenomena, from discrete clicks to a continuous noise, may be demonstrated by drawing the thumb nail at different rates along the teeth of a comb. The rate, however, must not be too slow, for in that case the perception of the interval between the qualities or events tends to intrude and to become the characteristic feature of the perception, and the succession as such becomes difficult to perceive.

If one takes his pencil and taps on a hard surface at first as rapidly as possible, and then slowly with about 1 sec. between taps, he will notice that with the rapid rate the sounds follow each other like beads on a string. It is impossible with this rate clearly to perceive an interval between the successive sounds, and the sounds themselves appear to belong together. It is just this belonging together of a series of discrete impressions within the conscious present that constitutes the essential character of the perception of succession. With the slow rate, on the other hand, the intervals between taps become focal and the taps serve as marks of the beginning and end of the interval between them.

The experimental study of succession consists in measuring in seconds or milliseconds the objective interval between stimuli which permits of succession. The important rates are those that determine the upper and lower limits of the perception. The upper limit is the shortest possible time that will yield the perception of succession. We shall call it the *threshold of discontinuity*. The magnitude of this threshold is, however, conditioned by various other factors, the most important of which is the kind of stimuli employed. Below are listed some representative results in milliseconds:

Hearing (electric sparks)	2.0 ms.
Hearing (falling weights)	5.6–6.7 ms.
Touch (impacts on fingers)	27.7 ms.
Touch (impacts on tip of finger)	55.0 ms.
Touch (impacts on other parts of hand)	45–150.0 ms.
Vision (at fovea)	22–61.0 ms.
Vision (at periphery of retina)	49.0 ms.
Vision (one stimulus at fovea, other at periphery)	76.0 ms.
Hearing (one stimulus to each ear)	64.0 ms.
Hearing (first) to vision (second)	60.0 ms.
Vision (first) to hearing (second)	71.0 ms.
Vision (first) to touch (second)	71.0 ms.

Examination of these results reveals the following trends: (1) Auditory succession is experienced with a much shorter

period between stimuli than tactual or visual succession. The latter have comparable limens. (2) In a single sense, temporal acuity may vary (*a*) with type of stimulus used, and (*b*) with the part of the sense organ stimulated. (3) When the two succeeding stimuli affect different parts of the same organ, or the first stimulus affects one sense and the second another, temporal acuity is lessened.

As the rate with which stimuli succeed each other becomes slower, the perception of succession becomes more difficult and tends to break up. When the time between the stimuli approaches about ½ to ¾ sec., perception of succession is replaced by perception of intervals bounded by stimuli. The rate at which the change occurs is influenced by the attitude of the observer and therefore is variable.

PERCEPTION OF TEMPORAL LENGTH

Duration as a length of time or of interval is perceived as a pattern within the conscious present. It is a length of time with a beginning and an end. Tap twice with your foot or pencil, leaving an interval of about 1 sec. between taps. The two sounds mark the beginning and the end of a definite stretch of time, and taken together with the interval constitute the pattern. The duration of the interval itself is the dominant characteristic, and the two taps serve as contours to mark the beginning and the end, much as an outline drawing on white paper separates the white space inside the figure from its background. We may also, of course, perceive a stretch of a continuous quality or event as, for example, a musical note. Here the beginning and the end of the sound itself are the contours of the pattern, much as the round edge or border of the full moon separates the disk from its background. Intervals of the first kind, those marked off by taps or flashes of light, are technically called 'empty time'; and intervals of the second kind, consisting of continuous sounds or pressures, are called 'filled time.' The so-called empty interval is, of course, not empty but has an indeterminate conscious con-

tent, quite different from the definite visual, auditory or tactual qualities of filled intervals.

All this is true of relatively short intervals, say of less than 5 sec. in length. We also experience the duration of much longer intervals, such as minutes, hours or days. Although the intervals of an hour, a week, or longer are not perceived directly, their duration, nevertheless, may be apprehended. With these very long intervals time perception is intellectual and indirect.

In most of the work on perception of long intervals, times of 10 sec. to a few minutes have been used. We may perceive such intervals as long or short, or as passing slowly or swiftly, but the integrated pattern of the perception is different from that of 'small' intervals. The interval, in either instance, has a beginning in some event and an end in another. So far, the patterns of long and short intervals are similar. The important difference is that the long interval is filled with a number of other events, many of which may have no relation to each other or to the 'end.' The experimental investigations of the perception of temporal length, therefore, fall into two groups according as they are directed to the 'short' or to the 'long' intervals.

The perception of short intervals. The principal method employed in the study of short intervals has been the method of comparison. In a single observation, two *stimulus intervals* are given and the observer is asked to say whether the second is perceived as longer than, shorter than, or equal to the first. The problem is to determine by psychophysical measurement the just noticeable difference or the differential limen, between two intervals.

The *length* of the interval employed determines to some degree the accuracy of time perception. Relatively small times tend to be perceived as greater (positive constant error), and relatively long times as smaller (negative constant error), than they actually are. The critical interval where the negative errors shift to positive, is characterized by little or no constant

error. This interval, which is perceived with maximum accuracy, is termed the *indifference point.* Above and below this point, errors of perception increase in size. The most frequently cited value for the indifference interval is approximately 0.75 sec.

An important determinant of the nature of errors in perceiving duration is the *attitude* of the observer. In reproducing short intervals, for example, variation in attitude may in some cases produce a reversal of constant error. Experiments reveal the need of much practice to produce relatively constant attitudes of observation. The reports of observers show, however, that even after practice it is very difficult to maintain a relatively constant attitude.

The conditions discussed above may be considered as general. Perception of short intervals is also influenced by a variety of special experimental conditions:

1. The perception of the temporal length of *filled* intervals is partly determined by quality of the stimulus. Tactually and visually filled intervals are perceived as longer than sounds of the same length, because of the persistence of the after-sensation in touch and vision.

2. A *filled* interval tends to be perceived as longer than an *empty* interval of the same length, whether the filling is from a continuous or an intermittent stimulation.

3. The *empty* interval, whose delimiting sounds possess the more striking characteristics, will be perceived as longer than an interval with indifferent boundaries. This difference is due to the fact that the apparent duration of the delimiting sounds is added to the interval itself, making a unitary pattern which is the basis of the perception.

The perception of long intervals. The investigations of the perception of long intervals have been made with intervals whose length varied from 8 or 10 sec. to about 200 sec. In addition to the method of comparison an alternative method has been employed which consists in the estimation of a duration in seconds. An observer, for example, may be required

to tap continuously with his pencil. After a certain length of time he is stopped and asked to state how many seconds he considers were in the interval covered by his tapping.

Several attempts have been made to determine the influence of variation in type of filling on the perception of time. It was found by the method of direct comparison that temporal perception depends less upon the number of separate presentations in the interval than upon the attitude of the observer toward them. The more the observer notices or becomes interested in the filling, the shorter the intervals are perceived to be. Words, for example, are shorter than noises, and meaningful sentences shorter than a series of nonsense syllables of the same objective length. Results obtained with the estimation technique reveal a similar tendency. With intervals of 15 to 30 sec. filled with tapping, and with cancellation of numbers, etc., temporal estimation again was influenced by the degree of interest or absorption in the task during that interval. The more engrossing the contents, the shorter was the interval judged to be.

The marked ability for 'time measuring' apparently possessed alike by certain idiots, normal people and animals has raised the question concerning its basis. Is there a physiological time sense? There is James's example of an idiot girl who screamed for her dinner at exactly the same time each day. Hall reported the case of an adult who could almost invariably give the time of day within a few minutes without reference to his watch, and apparently without noting the sun or other external conditions. This guessing was just as accurate on dark and cloudy nights as during the bright daylight. Other, similar examples have been reported. On the basis of these observations some writers infer that this accurate timing is made possible by a unique physiological mechanism, a 'time sense.' Such a conclusion is not valid. The observations cited were not controlled by elimination of environmental cues, and accuracy of guessing was frequently affected by practice. Furthermore the ability is apparently possessed by relatively

few people. Most of the persons reported upon have cultivated the guessing of time as a hobby.

Estimation of time after sleep is related to the time guessing discussed above. Many instances are reported of individuals who can waken very close to a predetermined time. This has been taken by some writers as an indication of an unconscious mechanism of temporal judgment during sleep. One investigator could, on the average, awaken within 10 min. of the set hour. Most of the awakenings came between 6 and 10 A.M. so that ample environmental cues were available. Since this method of awakening had been employed for a year prior to the investigation, the influence of practice seems obvious. In another experiment the subjects estimated the time and recorded the conscious basis of their estimates on being awakened between midnight and 5 A.M. The average error of their estimates was about 50 min. Cues which form the basis for this temporal estimation included degree of fatigue, degree of restlessness, sensations from stomach and bladder, closeness of thought to the topic in mind on retiring and amount of dreams recalled. It would appear that the temporal frame of reference in guessing time, either during the day or upon waking, is constituted of a great variety of auditory, visual and organic sensory processes supported by memory images. Practice in interpreting these cues, especially the extra-organic ones, apparently improves the accuracy of estimation.

The observation of animal reactions in certain situations reveals behavior characterized by temporal cycles which is analogous to perception of time. It is reported that fish in a stream adjacent to a railroad station assembled at the time scheduled for the arrival of the trains, when they were fed by the passengers. The fish made no response to trains arriving at other than the regular time. Certain animals, like hermit crabs and marine snails, show specific phototropic reactions that correspond in time to the oscillations of the tide. When these animals are removed to an aquarium or a laboratory, they continue to give the same reactions to light at the hours of

high and low tide although, of course, the stimulation from the tidal changes is entirely absent.

Various observers have noted that bees will come for honey at the specific hours when this food is available. Experiments with bees furnish good evidence for concluding that they possess temporal memory. This temporal memory operated day or night, and was independent of the alternation of light and dark and of weather changes. It appeared, however, to be bound up with a 24-hour rhythm.

In an isolated laboratory room a dog was given food regularly every thirtieth minute. After several repetitions a single feeding was omitted. At approximately the thirtieth minute, however, there occurred a secretion of saliva which had previously happened only when food was present.

These reactions of animals to time intervals are understandable in terms of *physiological rhythms.* Through many repetitions the reactions have become organized as part of the animals' physiological processes. When internal organic changes continue for the length of time which usually elapses between succeeding exposures to the rhythmic stimulation, the customary response appears in the absence of the original stimulus. The temporal interval has become part of an integrated pattern in the physiological rhythms of the animal.

In addition to the various conditions involved in the comparison and estimation of intervals which have been discussed, experiments have frequently revealed still other factors which seem to be involved in the perception. Muscular strains often occur which increase in intensity as the temporal stimulus continues and are replaced by relaxation when the stimulus stops. Again, expectancy that the end of the interval is about to be reached frequently arises. If the end comes before it is expected the interval seems short, if the end is delayed the interval seems long. Visual imagery also is sometimes found. For example, a line is seen whose length increases as the duration of the stimulus increases. Although some psychologists have been tempted to regard such factors, particularly tension-relaxation and expectation, as fundamental

to all perceptions of temporal length, the facts seem to be that they occur most frequently in the early stages of experimentation when, of course, they play their part in the total pattern. But in later stages they tend to disappear; the organism, to put it figuratively, no longer finds them necessary.

THE PERCEPTION OF RHYTHM

If, without counting, one gives attention to the movements of the body and limbs while dancing, walking rapidly or running, he will readily perceive a fairly consistent recurrent pattern of movements. Similarly the uniform puffs of a locomotive seem naturally to fall into rhythmical groups of two or of four puffs. Rhythm is the perception of groups or patterns of successive impressions, the members of which are perceived to be consistently different in some quantitative aspect, usually in relative intensity and duration. With this variation come subordination and synthesis of the elements into groups.

It was formerly thought that clearly defined rhythm was limited to auditory or auditory-kinesthetic impressions. Researches have demonstrated, however, that rhythmical perception occurs also in the tactual, motor and visual fields. Nevertheless auditory rhythm appears to arise most readily and to be the most convenient for investigation.

Repetition is essential for perception of rhythm. Although rhythm never arises from the presentation of a single rhythmical unit, it appears promptly with the first recurrence of that unit.

Rhythm also requires a periodic accentuation of certain elements in the series of sensory impressions. When a series of identical and equally intense stimuli is presented at a constant rate, and is attended to with a passive attitude, there appears a periodic change in the intensity of some of the sounds which divides the series into groups usually of two but sometimes of four sounds. The result is a *subjective accentuation* of the first member of each group. This subjective accentuation produces also an apparent lengthening of the in-

tervals between successive groups. Some observers experience subjective rhythm involuntarily or without instruction. For others it is readily induced by indirect suggestion. There are certain observers, however, for whom the rhythm is perceived only when they count and follow the stimuli with movements in some part of their body.

The number of stimuli integrated into subjective unit groups varies considerably. Groupings that may be classed as natural, easy or pleasurable have 2, 3, 4, 6 or 8 elements. The 2-group and the 4-group occur most frequently. The larger groups of 4, 6 and 8 elements tend to divide into subgroups of 2, 3 or 4 elements each. Although subjective accentuation may be altered to some degree voluntarily, odd groups of 5 to 7 are difficult either to obtain or to maintain without extended practice.

In a series of sensory impressions, any regularly recurrent impression, which is more intense or longer than the rest, subordinates the other impressions to it and grouping occurs. Such rhythm is *objective*. It is well illustrated by rhythm in dance music and in recited poetry.

To perceive rhythm the rate with which the stimuli recur must be neither too slow nor too fast. No subjective rhythm occurs with rates faster than about 8 per sec., or slower than about 1.5 sec. between stimuli. Ordinarily the 2-group appears first with the slow rates. As the speed is increased, the 3- and 4-groups appear, and then the 6- and 8-groups. If the rhythm is objectively determined, increased rate brings a combination of small groups into larger groups. At slow rates where rhythm just appears, the integration of the successive groups is weak. As the rate increases, the definition of the rhythmic form becomes more precise, the apparent intervals between succeeding groups grow larger and the accentuation within each group becomes more pronounced. Whenever the speed increases in this manner, the total time taken for any rhythmical group is, of course, shortened.

The maximum integration of successive groups that may be perceived as a whole has been termed the *temporal range*

of consciousness. Apprehension of the whole must occur, of course, without counting. A comparatively long series of impressions, such as equally spaced clicks, may be perceived as a united cluster if the rate of the succession is advantageous. One investigator found that the most favorable rate for apprehending audible strokes in groups was about 0.2 to 0.3 sec. Under these conditions as many as 40 strokes could be grasped as 5 subgroups of 8. This grouping yielded a maximum range of 12 sec. With the strokes occurring much slower, the longest duration of a rhythmical group was 36 sec. Other investigators, however, have obtained ranges extending from about 0.7 to 6 sec.

Rhythmical integration has a direct bearing upon musical and poetical practice. Small groups of impressions may recombine into larger groups. It is this kind of organization that produces musical phrases and poetical verses and makes possible their organization into still higher units. Neither the phrase nor the verse contains more than six measures, or feet, respectively. Furthermore the larger musical groups and the poetical stanza are seldom of more than four phrases or four verses. Thus they would seem ordinarily to lie within the range of apprehension.

We have already seen that the fundamental facts in the perception of rhythm are grouping and perception of the groups as wholes. Most explanations of rhythm have involved an analysis of the grouping experience. The recognition that the rhythm involves the perception of groups not given in the stimulus leads us to seek some psychological mechanism upon which this phenomenon can depend. In general a kinesthetic theory seems satisfactory. Observation reveals that the motor factor, the importance of which is emphasized by most investigators, is almost always present in rhythm. Apparently *kinesthesis is sufficient for the establishment of rhythmical perception.* In general the evidence indicates that perceiving rhythmical groups as wholes is due to sensations of movement or strain which accompany the qualities that are grouped. One makes a series of clicks into a rhythm by barely tapping

with his toe or barely nodding his head, or—as when the more overt movements are lacking—by accenting the clicks by ear-strain, *i.e.,* the kinesthesis of listening that arises in the muscles of the middle ear. The kinesthetic sensations are accompanied by a perceived change in the objective series of stimuli, and grouping results. The role of kinesthesis in rhythm may, however, vary considerably. It is, in general, most conspicuously associated with the first clear apprehension of the rhythmical form. Thereafter kinesthesis tends to drop out and the rhythm is carried by the visual, auditory or tactual pattern alone.

REFERENCES

1. Bolton, T. L. Rhythm. *Amer. J. Psychol.*, 1894, 6, 145-238.
2. Boring, E. G. *The physical dimensions of consciousness.* New York: D. Appleton-Century Co., 1933. Pp. 127-149.
3. Hall, G. S. The estimation of time. *Amer. J. Psychol.*, 1895, 7, 450-451.
4. James, W. *Principles of psychology.* New York: Henry Holt and Co., 1890. I, pp. 605-642.
5. MacDougall, R. The structure of simple rhythm forms. *Psychol. Monog.*, 1903, 4, 309-411.
6. Ruckmick, C. A. The rhythmical experience from the systematic point of view. *Amer. J. Psychol.*, 1927, 39, 356-366.
7. Woodrow, H. Individual differences in the reproduction of temporal intervals. *Amer. J. Psychol.*, 1933, 45, 271-281.

CHAPTER 11

THE PERCEPTION OF MOVEMENT

The astronomers say that the stars are continually changing their position at the rate of many thousands of miles per second. We, however, cannot see them moving. That is to say, there is no perception of their movement. Again, when we look fixedly at the minute hand of a watch we see no movement, but if, after a brief interval, we look at the watch once more we perceive the hand in a new position. In this case we perceive a change in position but not a continuous change. When we see an athlete doing a standing broad jump we see him, as we saw the hand of the watch, first in one place, then later at another place, but in addition we see him moving or changing his position from one place to the other. It is the perception of this continuous change in position of an object that is called the perception of movement.

Perceived movements have spatial properties. The fact that the object is seen at different places means that we perceive its localities. Furthermore the movement itself is localized; it occurs between the place of its beginning and of its end. Movements also have spatial extent; the distance through which the object travels may be perceived as short or long. Moreover, they have shapes; they may be curved, circular, spiral or serpentine. Finally, perceived movements have direction; the perceived object moves forward or backward, up or down, to the right or to the left hand, clockwise or counterclockwise.

There are also temporal properties in the perceived movement. Perceived movement has rate; it may be fast or slow. Movement is not only continuous; it may also be successive, as when we see a bird flitting from limb to limb. In this case

This chapter was written by Harry R. DeSilva of Massachusetts State College.

there is first movement, then rest, then movement again. And since the periods both of movement and of rest have durations of variable length, movements may change from successive to continuous as the rest period becomes progressively shorter. Finally, movements may also be grouped into rhythms as in walking and dancing.

GENERAL CONDITIONS OF THE PERCEPTION OF MOVEMENT

The fact that perceived movements have these spatial and temporal properties suggests that it is the definite and orderly *change* in the conditions which underlie the perceptions of distance, position, direction, size and shape of objects that brings about the perception of movement.

Of these changes, those in brightness, intensity, distinctness and size are especially significant for the perception of movement in the near-far direction. Thus we may perceive auditorially an approaching or receding locomotive by a continuous increase or decrease in the volume and intensity of its noise, or visually an approaching or receding automobile by the continuous change in its size and distinctness in the daytime and by the change in the brightness and size of its lights at night.

Successive stimulation. Change of another kind is necessary for perception of movement in directions other than approach-recession. It will be recalled that localization of objects occurs most frequently and is most certain in the visual field and on the skin. We are also able to perceive the posture of the body and the position of its movable parts—the fingers, arms, legs, head and eyes—with considerable accuracy. Furthermore it will be recalled that the receptors in the retina and in the skin are spatially distinct from one another and distributed as in a mosaic (see pp. 66f., 156f.). Since such a distribution is necessary, it is only in these sensory fields that movements are perceived in all directions. The change that conditions the perceptions of these movements is a *successive stimulation of the receptors* at a rate rapid enough to provide for the *con-*

tinuity of the movement. For example, when an insect crawls on the skin it successively stimulates pressure receptors. If the insect should fly across the field of vision the retinal image will move across the retina, and the cones that lie in its path are successively stimulated. When the forearm is moved by bending the arm at the elbow-joint, the surfaces in the joint rub against each other and the receptors are again stimulated successively. A certain (minimum) rate of stimulation is necessary, however, in order to obtain continuity of movement with the successive stimulation of discrete points.

The influence of bodily movement. When a person is in passive motion as, for example, when riding in a street-car or a merry-go-round, the situation becomes more complicated. The perception of stationary or moving objects is then determined by both retinal and kinesthetic stimulation. The somesthetic stimulus for the perception of the motion of one's own body, whether it be progressive or rotary, is an acceleration in the rate at which one is moving. Frequently, of course, there are accessory cues such as the jolting and swaying of the automobile in which one is riding. Normally the organism has no difficulty in apprehending which objects are moving and which are stationary, but, if some of the normal pattern of stimulation is changed, perception is changed accordingly. When, for example, one takes off in an airplane, he frequently perceives the earth receding from him instead of perceiving the change of altitude of his own body. Here, that part of the total stimulation which normally would determine the perception of his own motion is absent, and the retinal stimulation, in this case the continuous decrease in the size of the retinal images, determines the outcome. A similar situation occurs when one is seated in a train and first perceives his own car moving and then discovers that, instead, it is a train on an adjacent track that is moving. The moving train stimulates successive points on the retina, and this form of stimulation causes perception of movement. But this same retinal stimulation would occur if one's own train were moving and the other

train were standing still. Therefore, at first one is apt to perceive one's self as moving and only later, as a result of a slight change in the stimulus conditions such as the perception of a stationary object outside the coach, does the true perception occur. Here the stimuli were, like the reversible perspective figures, ambiguous. In active movements of the body such as walking, dancing, turning, twisting, etc., the cues for perception are so numerous that ambiguity of the stimulus conditions rarely occurs.

Influence of head and eye movements. If the conditions for perceiving the movement and stationary position of one's body are important for the determination of the object that moves, the conditions for the perception of movements of *parts of the body* are still more important. Especially is this true of the eyes. For when the eyes are in motion the retinal images of a moving object fall on a moving retina, and the displacements of a retinal image which is already moving must vary as the retina itself moves. Fortunately for perception, eye movements follow definite laws and only two kinds are of significance in the perception of movement. The first is *convergence and fixation* by means of which the retinal images of the object are brought on the foveas of the two eyes. The second kind is the *pursuit movements* by means of which the converged eyes follow a moving object.

The stimuli for the perception of eye movement derive from the turning of the eyeball in the fatty socket of the orbit, and, perhaps on account of the yielding nature of the tissues, the perceptions of eye movements are not as definite as those of arm and leg movements which rotate about the joints. Nevertheless we are able to perceive crudely the movements of convergence and of pursuit.

The integration of conditions. We have then a triple set of conditions for the perception of visual movement: the displacement of retinal images, and the kinesthetic stimulation from the moving body as a whole and from the head and eyes. At times every one of these various conditions plays its

due part in the perception. An instance is to be found in the experience of looking at a landscape from the window of a moving train. If an object in the middle distance is fixated it is perceived as stationary, but since the subject is himself moving, the eyes or head must make pursuit movements in order to maintain the fixation. All objects between the fixated object and the subject are perceived as moving in a direction opposite to that of the subject. All objects beyond the fixated object are perceived as moving in the same direction as the subject. This difference in direction is due to the fact that the retinal images of the near objects are moving across the retina in one direction, and those of the distant objects (those beyond the fixation point) are moving across the retina in the opposite direction.

At other times some one or a pair of the sets of conditions determine the perception of movement. Consider, for example, the following perceptions. When a bird flies across our field of vision and the eyes are stationary, the retina is successively stimulated at different places (retinal images) and we see the bird move. But if the bird is stationary, the head and eyes may be moved in such a way that the same retinal places are stimulated as were stimulated by the flying bird. Nevertheless the bird is usually perceived as stationary (movements of the head and eyes). Furthermore the eye may follow a moving object so that only one place on the retina is stimulated and yet the object may be seen as moving (movement of the eyes overcomes the retinal image). Or, finally, if while fixating a stationary object the position of the eyeball is displaced by pressure upon it with the finger, the object is seen to move (displacement of the retinal image). In all these perceptions both sets are always functioning, but the balance between them which derives from the total situation of the moment determines the outcome.

The attitude of the subject. The situation in which the perception occurs always includes the subject himself, and there is no question but that his past experience has molded

his perceiving in such a way as to enhance the effect of some stimulus conditions and to inhibit that of others. We expect birds and airplanes when seen against the sky to be moving; we expect mountains and trees to be stationary. When we know that the car in which we ride is moving, the perception of roadside objects as stationary is thereby enhanced. Against a background of sky and moving clouds the moon may be perceived as moving and the clouds as relatively stationary, but by a shift in attitude the moon may become stationary and the clouds moving.

After-images of movement. If we look fixedly for fifteen or twenty seconds at a slowly rotating white disk upon which has been drawn a heavy black spiral line, the disk will seem to contract or expand depending upon the direction of rotation of the spiral. If we then turn our eyes away and fixate a person's face, the face will appear to expand or shrink depending upon the direction of rotation of the spiral. Similarly, if we look at a waterfall, or at a flowing stream, for a minute or so and then glance away at the landscape, the latter will appear to move slightly in the opposite direction. Or if we are riding on the rear platform of a train, and it suddenly stops, we notice that the formerly shrinking and receding objects appear to broaden or come nearer. In short, under certain conditions a movement produces an after-effect which manifests itself as a movement in the opposite direction. The velocity of the after-image of movement corresponds roughly but by no means exactly to the relative velocity of the moving object.

EXPERIMENTAL INVESTIGATION OF PERCEIVED MOVEMENT

Thus far we have been concerned with the *nature* of the perception of movement and its *general* conditions. We now turn to the more special results of the experimental investigations of the perception. The experiments may perhaps best be classified according to the nature of the stimuli employed.

Movement with a moving stimulus. One of the earliest experimental problems in this field was the determination of the minimal distance that the stimulus must travel for movement to be perceived. This distance is conditioned in part by the rate at which the stimulus travels. It depends also upon other conditions such as the nature of the illumination, the nature of the background and the part of the retina or skin that is stimulated. The particular values found have little meaning, therefore, unless the special conditions are known. Three things may, however, be said of them. (1) They are always larger than the extent of the involuntary tremors of the eye that occur whenever we try to keep the eyes still. (2) Except on the fovea, they are always smaller than the two-point limen. Since this limen is taken as the measure of visual acuity and of cutaneous sensitivity, movements on the periphery of the retina and on the skin may be perceived distinctly when objects are indistinct. (3) The magnitude of the threshold is smaller on the more sensitive parts of the retina (the fovea), and of the cutaneous surface (the finger tips or the lips). On the periphery of the retina the decrease in sensitivity to movement is not proportional to the decrease in visual acuity. Thus at the center of the retina the two limens are approximately the same, but at a distance of 20° from the center the limen for acuity is four times larger than that for movement. For example, when we look directly at a moving automobile we can see the automobile as distinctly as we can see its movement. When we look out of the corner of our eyes we cannot see either one as well as we could in direct vision, but we can see the *movement* of the automobile before we can see the automobile *itself*.

Another problem has been the determination of the limiting rates of stimulation that will arouse movement. In vision under optimal conditions and at a distance of 2 meters the minimal threshold is about 0.2 cm. per sec. The minute hand of a watch would have to move some five or six times faster than it does, to be seen as moving at 2 meters' distance. There is also an upper threshold, for if the stimulus moves too rapidly

the object will be perceived as a flicker or a blur. The magnitude of this threshold is also variable according to the conditions. One laboratory result was 150 cm. per sec. when the moving object was 2 meters away. From the point of view of our everyday experiences this rate seems slow, for it is only about 30 miles per hour, and we can perceive moving automobiles at several times this rate. If, however, a moving automobile were only 2 meters away and were seen through a small window, it is questionable whether we should perceive it as moving.

The distance away of a moving object decreases the *perceived rate of movement.* An automobile traveling at 60 miles per hour will appear to be moving rapidly at a distance of a few yards but slowly at a distance of a few miles. A ship on the horizon is not perceived as moving at all. The decrease in rate is, however, not proportional to the distance. The tendency for objects to maintain their size with increasing distance (size constancy, see p. 215) tends to keep constant the apparent rate as distance increases. For example, if one backs away from a revolving barber's sign it will not appear to decrease in rate of rotation as much as it would if it decreased proportionately in perceived size.

Movement with a single stationary stimulus. It was the astronomer von Humboldt who in 1799 first observed that a single star in the dark sky appears to move about when steadily fixated. Many years later this same phenomenon was produced in the laboratory and is called the *autokinetic sensation.* If a small spot of light is fixated in a dark room, it tends after a few seconds to move irregularly, usually toward the periphery and upwards at an angular velocity of from 2° to 10° per sec. The extent of the movement is seldom more than 40°. The movement does not seem to be conditioned upon eye movements because it occurs when the spot is carefully fixated. There is no accepted explanation for this effect, although one experimenter holds that it is best accounted for in terms of the streaming of bioluminous fluids in the eye-

ball itself. The one condition is that the observer must have no objects in the visual field with which to localize the point of light. This lack of an object of reference undoubtedly leads to unsteadiness and uncertainty in localization.

Stroboscopic movement. Since the perception of movement results from the successive stimulation of different receptors, there would seem to be no fundamental necessity for the stimulus itself to be moving. A series of *discontinuous stimuli* that would successively stimulate different receptors should also produce the perception of movement, provided the rate of succession were adequate. This is exactly what happens when we see the movies move. The stimuli in the cinema are a series of photographs projected on the screen. Every one of the photographs represents a slightly different static picture of an object that was moving when the pictures were taken. When the series is projected on the screen in the proper order, and at the proper rate, usually 24 per second, normal movement is perceived. The quality of the movement depends upon the rate of projection. If this rate is very slow, one sees a succession of static pictures. As the rate is increased the movement becomes first a flicker, then normal and finally jerky and blurry. Beyond this stage all movement disappears and one has only an impression of a filmy surface, such as one sees in looking at a rapidly revolving color wheel or electric fan.

One of the disadvantages of having an upper threshold for the perception of movement is that we cannot observe the details of many ordinary everyday happenings such as how a golf ball responds to a blow, how a piece of china breaks or how a sleight-of-hand performer fools us. This human limitation is now being circumvented by the use of a mercury-vapor-tube stroboscope. Used in a dark room this lamp illuminates a moving object such as a moving golf club several hundred times per second, so that a very rapid moving-picture camera takes still pictures of various stages of the rapid movement. After development the film is run at slow rate

so one can observe the smallest details of how the golf ball is deformed when it is hit and how it regains its roundness as it moves away.

Experimental work on the largest amount of angular extent that may be permitted between separate successive pictures on a moving-picture film has shown that continuous motion is disturbed at all frequencies for distances greater than 4.5 angular degrees and that no motion of any kind is possible above 12.5°. This possibility of obtaining an impression of uniform motion with angular separation of successive phases as great as 4.5° indicates that the mechanism which determines the continuity of visual motion is probably quite distinct from that underlying acuity, since the threshold of acuity is much less (see p. 215).

Movement with two successive stationary stimuli at different places. If a vertical line (*A* in Fig. 95) is presented and then immediately after an interval of 60 ms. a horizontal line *B* is presented, the vertical line will be seen to rotate clockwise from the 12 o'clock to the 3 o'clock position as in *C* of the figure. If the interval of exposure of the two lines is shortened to about 20 ms., the two lines appear simultaneously and form a right angle. If, on the other hand, the interval between the exposures is lengthened to 200 ms., the two lines will be seen successively with no movement at all.

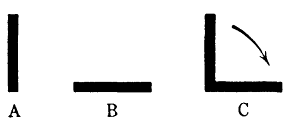


FIG. 95. STIMULI FOR APPARENT MOVEMENT

This kind of movement, often known as 'apparent movement' because it is elicited from a non-moving stimulus, has received an extensive investigation in the laboratory. One reason for its popularity is that an important trend in modern psychology (Gestalt psychology) took its rise from a study of the perception of movement under these experimental conditions. Another reason for the large amount of study undertaken on this phase of movement is that it has furnished facts

upon which to build neurological theories to explain visual perception. A form of apparent movement when the stimuli occur very rapidly, *e.g.*, in times of the order of 60 ms., is one in which no object is seen to move. The perception is readily identified as movement and is not readily described in analytical terms. It has been given a special name, the 'phi-phenomenon.' The important variable conditions upon which this form of visual movement depends are the following: (1) the length of the pause between the first and second stimulus, (2) the intensity of the stimuli, (3) the distance between them, (4) the exposure time of each, (5) the magnitude of the stimuli and (6) the attitude of the observer.

If two vertical slits in a black cardboard are successively illuminated to produce apparent movement, a certain range of light intensities is most favorable for eliciting the phi-phenomenon. A great reduction of illumination results in poor movement, whereas a great increase in intensity results in each figure first expanding ('gamma movement'). If the light intensity of the second member is made much stronger than that of the first member, the movement sometimes reverses itself, passing from the second brighter member back to the first member and then back to the second again ('delta movement').

It has also been shown that, when the light intensity is increased, the resulting change in movement can be compensated for by decreasing the pause between exposures, or by increasing the distance between the stimuli. If any one of the three conditions—(1) light intensity, (2) pause and (3) distance between stimuli—is varied so as to lessen the movement, a reciprocal change in one or both of the other variables within certain limits restores the movement. These relationships are, however, limited in scope since there are many other significant conditions. Thus, as the experienced motion-picture engineer also finds, besides these factors the size, texture and reflection coefficient of the screen, the general illumination, the background, the distance and angle of the spectator from the

screen must be considered in order to produce satisfactory moving pictures.

Whatever the other conditions, exposing two lines a short distance apart always produces a better perception of movement than exposing them farther apart. The motion-picture cartoonist recognizes this fact and makes sure that the difference between his successive drawings is very slight. The less the displacement between successive drawings the more lifelike and complete the movements of the figures.

Exposing both stimuli briefly gives an impression of swifter movement, whereas lengthening the exposure time slows down the movement and makes it jerky. Lengthening the exposure time of the first stimulus only makes the movement more jerky and sometimes produces a reverse movement. If the pause is decreased, the time of the first member must be relatively long. In this case the second stimulus can be exposed before the first has disappeared and yet good movement will be seen. Conversely, long exposure of the second stimulus injures good movement.

In general, the smaller the stimuli the better the movement; the larger the stimuli the jerkier and less satisfactory the movement. Thus we reject the front seats in the movies to avoid jerkiness and flicker, and likewise shun the rear seats in a large theatre so as to be sure to see the desired details.

The *attitude of the observer* often determines the direction of the movement. If a subject observes successively appearing lines going from left to right for a number of presentations, and then, if suddenly the direction of the stimulus movement is reversed without his knowledge, the subject will continue to see the lines moving from left to right for three or four presentations. If dots in this position ˙ . are followed by dots in this position . ˙ the observers may get a right-left, left-right movement, or an up-down, down-up movement, or a clockwise or counterclockwise spinning movement. Many an observer can at will, by changing his point of regard or by self-instruction, get any one of these three different kinds of movement.

If one point is applied to the skin, and shortly after another point is applied at a neighboring place, an observer sometimes gets the impression of tactual movement from the first to the second position. The best impressions occur when the interval is about 75 ms. In about half the cases, however, even with the optimal times and distances, observers do not get a good experience of movement. *Apparent tactual movement* is not nearly so compulsory as apparent visual movement.

One of the most interesting forms of tactual movement is a 'bow movement,' in which the moving impression jumps into space above the skin from the place where the first stimulus was localized to the place where the second stimulus is localized. Some observers also report that the tactual movement is perceived in whole or in part by visual images, that is to say, they see rather than feel the tactual movement.

THEORY OF PERCEIVED MOVEMENT

The perception of movement occurs, as we have seen, under a wide variety of circumstances. In so far as the perception involves locality, its conditions are those of localization in general and need not be further discussed here (see pp. 228f.). In so far as the perception involves continuous change in position, its conditions are those of spatial continuity at large. The receptors in both the retina and the skin are punctiform with spaces between them. A moving stimulus excites these receptors successively, but, if the rate of this succession is adequate, we perceive, not a succession of impressions in different places, but a continuous change in position. The experiments in apparent movement show that the receptors may be widely separated and still, if the other conditions are realized, movement results. Apparent movement is, therefore, a limiting case of the perception of movement at large. The problem of movement is the problem of how perceived continuity issues out of discreteness in the stimulus.

It appears that movement comes about in this case because the second stimulus is substituted in some physiological man-

ner for the first. If the excitation aroused by the first could 'flow over' in some fashion so as to become the excitation of the second, continuity might be expected to result. A theory of a cortical short-circuiting of excitation has been proposed along these lines, and another theory takes account of the fact that the cross-flow of energy in the brain must be such as to reestablish the form of the moving object. These theories, however, are highly speculative in their physiological assumptions and cannot be accepted literally. Nevertheless they point to the generality basic to any theory of movement: Movement must emerge from successive discrete stimulation by way of a rapid integration of two or more spatially separate excitatory processes into a dynamic continuum. Only in this way can we explain how movement can be seen when there is no moving stimulus. Before discovering where and how this integration occurs physiologically we shall have to answer many puzzling questions about the dynamics of perception.

REFERENCES

1. Brown, J. F. The visual perception of velocity. *Psychol. Forsch.*, 1931, 14, 199-232.
2. DeSilva, H. R. An analysis of the visual perception of movement. *Brit. J. Psychol.*, 1928, 1, 550-575.
3. Dimmick, F. L. An experimental study of visual movement and the phi-phenomenon. *Amer. J. Psychol.*, 1920, 31, 317-332.
4. Guilford, J. P. Autokinesis and the streaming phenomenon. *Amer. J. Psychol.*, 1928, 40, 401-417.
5. Higginson, G. D. The visual apprehension of movement under successive retinal excitations. *Amer. J. Psychol.*, 1926, 37, 63-115.
6. Hulin, W. S. An experimental study of apparent tactual movement. *J. Exper. Psychol.*, 1927, 10, 293-320.
7. Miles, W. R. Visual illusions of motion. *Sci. Mo.*, 1928, 27, 481-491.
8. Wohlgemuth, A. On the after-effect of seen movement. *Brit. J. Psychol. Monog.*, 1911, No. 1.

CHAPTER 12

PERCEIVING

Perceiving is a fundamental activity by which an organism gauges the current turn of events and their implications, both in the organism's surroundings and within its structural limits. Through this activity the organism becomes aware of internal changes as well as of external incidents. From time to time the organism surveys itself, or apprehends the appearance of other objects. It also perceives the properties of objects and events—their parts, groupings and interrelations. By this varied perceiving, we are primed to act, remember, imagine, comprehend and think. These other activities may then serve, in their turn, to extend our envisagement of ourselves and of the world.

Occasionally, perceiving turns out certain products that seem odd. We may, for example, perceive bare movement without perceiving the moving object, an odor or a sound without identifying the odorous substance or the source of the sound. These objectless products of perceiving, as we may call them, are as significant for an understanding of that activity as the more familiar objects and events which it brings before us.

It is important to realize that objects, events and their properties *as we perceive them* are the outcome or result of definite bodily processes. The perception of snow, for example, involves (1) retinal stimulation, (2) possibly also the stimulation of other receptors, (3) transformation of the stimulus energy before it reaches the receptors and also in the receptors, (4) release and propagation of neural impulses and (5) regrouping and repatterning of the neural impulses at various stages of their journey in the neural system. In order fully to under-

This chapter was written by S. Feldman and H. P. Weld of Cornell University.

stand the perception of an object, we must relate the perception to every one of these stages of bodily activity. At the present state of our knowledge we can, however, do this only in part.

The bodily activities and structures involved in perceiving undergo development in the individual organism from the embryonic period onward. Consequently perceiving is different in the infant and in the more mature individual. Furthermore, differences in perceiving among various orders of animal life point to the long evolutionary history of this activity. A knowledge of its history, both individual and racial, helps us to understand perceiving in the adult human organism. We shall, however, consider only certain aspects of its development in the lifetime of the individual.

CONSTANCY OF PERCEIVING

Whenever we see blood we see it as red. We always see coal as black, snow as white and ashes as gray. We perceive these objects in this *constant* manner even when we regard them from a distance or under different illuminations. The distant snow seen at night through the window may reflect no more light into the eye than the piece of coal lying in the firelight. Yet, even under these circumstances, the snow looks white and the coal looks black. This means that our perceiving is relatively constant in spite of certain changes in the stimulus, and this constancy is so familiar that it never occurs to us that it might easily be otherwise.

We perceive objects as constant in other respects than in color and brightness. We perceive them, under a variety of stimulus conditions, as constant in size (see p. 214), in shape, in weight, in temperature and in still other properties. In fact, much of our perceiving is thus standardized.

This constancy of perceiving depends in part upon the relative effectiveness of different aspects of the stimulus. The influence of stimulus factors that leads to a constant perceived resultant has been investigated chiefly in the perception of brightness.

Brightness constancy and general illumination. An object, a wall for example, that is seen as white in sunlight is also seen as white in moonlight. That is to say, the object appears the same under different illuminations. This fact becomes puzzling as soon as we realize that the amounts of light reflected into the eye by the object differ enormously. How great the difference is appears from the estimate that sunlight is 800,000 times as strong as full moonlight.

If, instead of being constant, the brightness of objects varied directly with the amount of light they reflect, we should find it very confusing. The white wall, for example, would look black in moonlight. Up to noon, when sunlight is most intense, all objects would actually grow whiter; and from noon on, they would gradually turn blacker. Similarly the paper of this book would look white near the window and dark gray at the other end of the room.

The problem of brightness constancy can be illustrated in yet another way. Under any illumination, the best white paper reflects only 60 times as much light as the deepest black paper. Consequently, by placing the black paper in an illumination that is 60 times greater than the illumination of the white paper, one can make the two surfaces—black and white—reflect into the eye the *same amount of light* and thus stimulate the retina with equal intensity. Nevertheless, we know that the two surfaces will not look alike. The black surface will be seen as black in a high illumination, the other surface as white in a lower illumination. The black surface retains its characteristic constant brightness in spite of the change in illumination.

It is possible to explain this constancy on the basis of familiarity. That is to say, since we know that the object is black we continue to see it as black. Often, it is true, specific acquaintance with an object does contribute to the constancy of perceiving. Such acquaintance or familiarity, however, can be shown to play no part in the constancy of our highly illuminated black. The white and black papers can be selected and presented to the observer without giving him

any opportunity to know beforehand what he is to be shown. The observer, nevertheless, sees the highly illuminated black as black.

What we find in this case of brightness constancy is a peculiar connection between *illumination* and the perception of *surface brightness*. Increasing the illumination of the black paper affects the retina in two ways. The increase in illumination raises (1) the intensity of the local stimulus (the amount of light reflected into the eye from the object alone) and also (2) that of the total stimulus (the amount of light reflected from the sky and all the objects about). The former effect, if left to itself, would tend toward a lightening of the black. *This tendency, however, is held in check by the rising intensity of the total stimulus.* As a result, the black remains relatively unchanged. Similarly, the *darkening* of the white wall seen in the moonlight is checked by the *low* intensity of the total retinal stimulus.

How this relative influence of stimulus factors is effected can be learned directly only by a study of the integration of neural impulses at various levels of the neural arc. The nervous system is already known to possess regulative mechanisms that control the constancy of the physiological functions. Some of these mechanisms, for example, are responsible for the maintenance of a constant body temperature in warm-blooded organisms such as ourselves.

The perception of space and of illumination. Besides regulating the effect of the local stimulus, the total retinal stimulus has independent effects of its own.

It is responsible, in the first place, for our perception of space. Objects are always seen as being in space. One may, however, perceive space without objects. For example, when we close our eyes or enter a lightless room, we say that we see nothing. What we mean is that we see no objects. Only the truly blind, who lack retinal stimulation altogether, see literally nothing at any time. Normally, the retina is active even in the absence of light, for there is a source of stimula-

tion in the eye itself (see pp. 74f.). *It is because local differences of stimulation are lacking that one does not see objects in the dark.* One does, however, see a dark, contourless field—an objectless space. This seemingly futile form of perceiving depends upon the uniform stimulation of the retina as a whole.

The total retinal stimulation is responsible, in the second place, for our perception of the *general illumination.* It enables us to perceive the prevailing *state* of illumination as well as changes in the *intensity* and *quality* of illumination. In the case of the illuminated black paper, the total stimulus enables us to perceive the high illumination at the same time that it regulates our perception of the object itself. Brightness constancy and perception of illumination are thus related.

We sometimes perceive illumination without illuminated objects. In a certain disease of the eye, the light entering the pupil, instead of being concentrated by the lens so as to form a sharp image, scatters over the entire retina. The visual field is consequently always a diffuse bright or dark mass. The brightness of the mass varies, of course, with the intensity of the light entering the eye, so that a perception of the prevailing illumination is possible.

The visual field takes on this character to the normal eye when a person is surrounded by heavy mist, looks out from the window of a bathysphere or enters a chamber cut inside a glacier. The walls of the glacier transmit into the eye a uniform blue light. Since stimulation is homogeneous, the person inside the glacier does not see the walls at all but only a formless field of blue.

Brightness constancy and special illumination. An object is usually seen in a general illumination. It may, however, have in addition a special illumination of its own. For example, it has a special illumination when a spotlight or when a shadow is thrown on it. The former would tend to increase, the latter to decrease, the brightness of the object.

We have seen that brightness constancy and the perception

of general illumination are related in that both depend upon the total retinal stimulus. That an analogous relationship obtains between brightness constancy and the perception of *special illumination* has been demonstrated in a number of experiments.

A simple experiment of this kind consists in casting a small shadow on a white card. On examining the card one of course sees the shadow and observes also that the part of the card where the shadow is looks a little darker than the rest of the card. Suppose now that a wide black line is drawn over the fuzzy edge of the shadow. There is a startling change. The shadow is no longer seen. In its place one gets the impression of a *very dark stain.*

Obviously, casting the shadow on a part of the card reduced the amount of light reflected from that part. Nevertheless, that part of the card remained approximately constant in brightness as long as the shadow was visible as such. It immediately turned darker, however, as soon as the perception of the shadow was abolished by the black line. The approximate constancy of the brightness broke down.

This experiment shows, then, that our perception of brightness is influenced by our perception of shadows in the same way that it was earlier shown to be influenced by our perception of general illumination. And the explanation lies again in the relative influence of different aspects of the stimulus. The fuzzy edge is the aspect of the total stimulus that is responsible for the perception of shadow. It also influences the perception of the brightness of the shadowed part of the card, because the instant the fuzzy edge is replaced by the black line the brightness changes.

This relation between the perception of shadow and the constancy of brightness is further demonstrated in the following experiments. By suspending a circular object between a white disk and the source of illumination, it is possible to cast a shadow on the disk so that *shadow and disk coincide.* Such a fitting of the shadow to the disk causes the sharp contour of the disk to prevail over the soft contour of the shadow.

The white disk turns dark, while the shadow, as such, becomes invisible. If, however, the shadow-casting object is moved slightly to one side so that the shadow no longer coincides precisely with the disk, the shadow becomes visible, and the entire disk resumes its normal brightness.

A black disk may, in turn, be lightened by a projection lantern in such a way that *the cone of light coincides with the circular area of the disk*. The black appearance of the disk may then be restored by moving the lantern so that the light no longer exactly coincides with the area of the disk.

Whenever, then, we see an object in shadow, our perception of the brightness of the object does not depend solely upon the amount of light that is reflected from the object surface, for, if this were true, we should see a shadowed white as much darker than we do. Similarly, when we examine an object in the light of a desk-lamp, its brightness does not correspond to the amount of reflection. Otherwise, a black object would look almost white. In both cases our perception of brightness is influenced by our perception of the object's special illumination.

Another method for eliminating the perception of shadow and thus also destroying the constancy of brightness, makes use of a screen with a hole in it or, as it is usually called, a reduction screen. The screen is placed in front of a shadowed object so that only part of the middle of the object is visible through the hole in the screen. Under these circumstances, the shadow is not perceived. At the same time constancy breaks down. If the object behind the screen is a shadowed white object, it now looks dark.

The reduction screen is experimentally useful because it enables us to *measure* the degree to which constancy prevails when the shadow is visible.

In a typical experiment, the observer (*O*) faces two disks (*A* and *B* in Fig. 96, I) that are illuminated from a window (*W*). Between the disks is a screen (*S*) that casts a shadow on disk *A*. Disk *A* is white, and is seen as a white surface in shadow. On disk *B* (the comparison disk) are sectors of

white and black. The sectors on this disk are adjusted so that, when rotated, it will present to the observer an *unshadowed white* that matches the *shadowed white* of disk *A*. Now if constancy were perfect, the shadowed disk *A* would be matched with 360° of white on the comparison disk *B*.

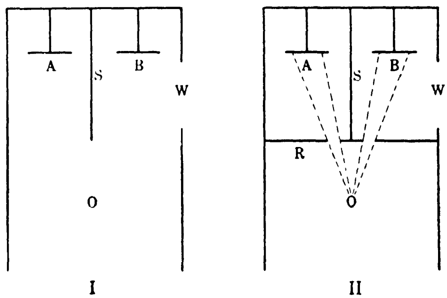


FIG. 96. PLAN OF AN EXPERIMENT IN BRIGHTNESS CONSTANCY.

As seen from above. The meaning of the symbols is given in the text. In No I there is no reduction screen. The dotted lines in No. II show the parts of the disks that can be seen through the holes in the reduction screen (R).
After Katz.

Actually the shadowed disk looks a little dark, so that the match on the comparison disk contains a sector of black in addition to the white sector.

The first column of Table XIII, reading down, shows the size of the *white* sectors on *B* (the comparison disk) with increasing depth of shadow cast by screens (*S*) differing in

TABLE XIII

I	II
246 0	202 4
180 0	80 1
178 6	65 9
116 0	4 2

transparency. The figures in this first column show that, as screens of progressively less transparency are used, the shadowed surface on disk *A* becomes progressively darker, *i.e.*, it takes less and less of white to match the shadowed disk. This

is what we should expect and as yet there is no evidence of constancy. We have reason, however, to suspect constancy in some degree, and in order to discover and demonstrate it we must change the experiment.

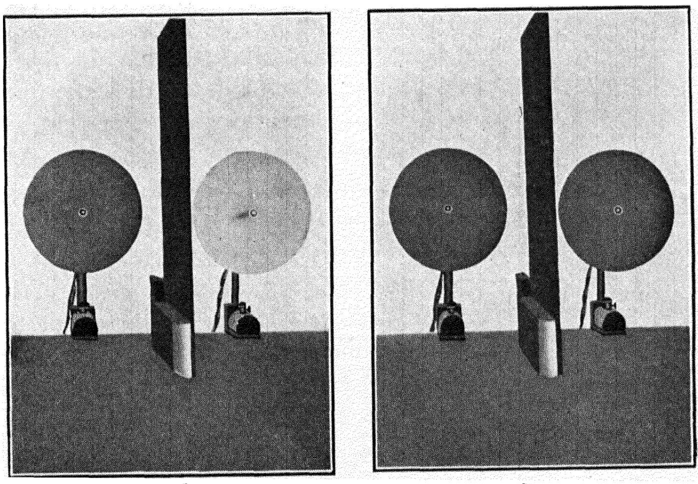
To determine the degree of constancy, then, a reduction screen (*R*) is placed in front of the two disks. This screen has *two* holes, and these are so placed that part of each disk is seen through each opening (Fig. 96, II). When disk *A* (the shadowed white) is now observed through the opening, one no longer, of course, sees the shadow, and the disk looks much darker. This change in brightness is confirmed by again matching the two disks. The results of these new equations obtained with the use of the reduction screen are given in the second column of Table XIII. The figures in this column represent the brightness value of the disk in exact correspondence with the stimulus intensity. The difference between the figures in the two columns gives us a measure of the degree of constancy. For instance, consider the difference between the two top figures of the two columns. It takes 43.6° more of white to match the shadowed disk without the reduction screen than with it. That is to say, disk *A* looked that much whiter than it would if its brightness were solely a function of the light reflected into the eye; consequently, 43.6° is a measure of the degree of constancy.

Fig. 97 shows two such matches photographed. The photograph demonstrates that, when the reduction screen is used (part II of the figure), perception is governed solely by the amount of light reflected from the disk surfaces. Consequently the matched disks also match for the camera. Without the reduction screen the matched disks do not match for the camera (see part I). Whereas the functioning of the camera is always strictly determined by the amount of light admitted into it, perceiving is in this second case determined only *in part* by the amount of light admitted into the eye.

The change from vision that is approximately constant to the 'reduced vision' resulting from the use of the reduction screen

is something that a painter learns to do at will. "Every painter represents a white object in shadow by means of gray pigment, and, if he has correctly imitated nature, it appears pure white."

Colored illumination. So far we have considered the relations of surface brightness to colorless illumination. A certain degree of constancy has also been demonstrated in



I II

FIG. 97. TWO PHOTOGRAPHS SHOWING BRIGHTNESS CONSTANCY.

The light comes from a window (not shown) on the right. The disks on the left in both pictures are white, but one stands in a shadow cast by the vertical screen. Each of the disks on the right contains sectors of black and white adjusted to match the shadowed surface of the left-hand disk. No. I shows the match made without, and No. II the match made with the reduction screen.

After D. Katz. *Die Erscheinungsweisen der Farben.* Leipzig: Barth, 1911, p. 181.

connection with colored illumination. A white disk appears tinged with red when it is illuminated with red light. The reddish tinge, however, is not so pronounced in ordinary observation as it is when the disk is observed through the hole of a reduction screen. That is to say, the white disk normally tends to retain its colorless character despite the colored illumination. The same resistance to change may be demonstrated for *colored objects.* A green disk in red illumination looks

practically colorless behind the reduction screen. In ordinary observation it retains much of its greenish appearance.

Other instances of constancy. The perceived size and brightness of distant objects are held constant by the governing influence of the perception of distance (see pp. 213f.). In the same way, perceiving the orientation of one's eyes and body makes for constancy in the perceived position and shape of objects, in spite of the fact that every change in orientation shifts and distorts the retinal image (see p. 229).

Film-color. The reducing screen does more than alter the brightness of an object. It transforms the product of perceiving much more radically. When we ordinarily inspect an object —say, a polished piece of wood—the outcome of our perceiving is rich in scope. We perceive the *grain* and *polish* of the wood, its *hardness* and its *weight.* We can tell whether the wood retains its *natural color* or whether it has been stained. We see it in a definite *place* and as having a definite *orientation.* We can perceive the *shadows* clinging to its surface and the *lights* reflected from its interior. When the same piece of wood is observed through the opening in the reduction screen, *all* the products of perceiving just listed *drop out.* Instead, one sees an objectless plane of color with a soft, sky-like texture. This color plane has no resemblance to the surface of a solid object. In fact, the observer has no way of telling whether he is looking at a piece of wood or at some other object. The plane is also very difficult to place, and it *always faces the observer directly* even if the object behind the screen is actually tilted.

The peculiar turn that perceiving takes in the case of this *film-color,* as it has been named (see pp. 58f.), shows how intricately the conditions of perceiving are interrelated. If one crucial stimulus condition is eliminated, perceiving may suffer a loss not in one respect alone but in a variety of ways. Obviously, such mutual regulation of perceiving by several aspects of stimulus implies definite neural activities of a very intri-

cate sort. And, in fact, certain brain injuries affect perceiving in much the same way as the reduction screen does.

One such case of brain injury is on record. The injury affected perceiving in many directions, both by way of disturbance and complete loss. Color vision was gone, and so was *surface vision.* The patient saw all object surfaces as soft, thick, spongy layers. Dark surfaces looked thicker than light surfaces, so that a series of grays from white to black appeared to him as *advancing stepwise.* He saw the series as lying in the same plane only when the darker grays were suitably pushed back. Furthermore, when he touched an object, his fingers seemed to him to be dipped in a transparent mass of color, whose thickness was estimated to be 6 to 8 in. when the surface was black, and 1 to 2 in. when the surface was white.

Some of these phenomena can be observed also in normal film-vision. The film-color seen through the hole in the reduction screen is not always a plane. It is very labile and instable. As one looks at it, it may begin to expand into a contourless mass; it takes on the appearance of a colored portion of space. It may even advance beyond the opening and present a surface lying *in front* of the screen. A finger or pencil inserted into the opening is seen as immersed in a transparent body of color.

Just as the perception of an object is radically transformed by merely covering up its contour and its immediate surroundings, so a simple change of another sort may restore the normal perception. If a sound is produced by tapping the object behind the screen, or if the observer is allowed to touch the surface of the object with a stick, the object not only acquires a *stable position* but also recovers many of its other distinctive properties. That aspect of the auditory or somesthetic stimulus which is usually regarded only as a condition of distance is thus shown to be a part-condition of many other properties of objects.

The use of the reduction screen represents one means of unraveling the many intricate conditions that are involved

even in one's routine perception of objects. Our principal concern, however, has been with its use as a way of abolishing the perception of *special illumination*—shadows and high lights. It is obvious that the reduction screen does not abolish the perception of *general* illumination. The film-color behind the screen is perceived, like the screen itself, *in the general illumination of the room.* The reduction screen cannot, therefore, destroy constancy altogether. Constancy is destroyed only to the extent that it depends upon the perception of special illumination.

INSTABILITY OF PERCEIVING

Perceiving is an event that takes time even when it appears to be instantaneous. The lower limit of the time of perceiving may be determined approximately by measuring the time elapsing between the presentation of a stimulus and the giving of a prearranged signal by the perceiving observer.

There is an upper limit as well. We have seen how perceiving may be constant despite certain changes in the stimulus. In time, however, perceiving may alter even when the stimulus remains the same.

Adaptation. We are already familiar with the facts of qualitative adaptation. Colors, odors, tastes and cutaneous qualities undergo characteristic change in time. The cause of these changes is being sought in the receiving mechanisms.

Many of the *spatial* products of perceiving alter in a manner similar to sensory adaptation. If one looks in a magnifying mirror, the Brobdingnagian face is perceived as *diminishing in size.* If one then looks in an ordinary mirror, the face seems Lilliputian at first. Such negative after-effects are especially pronounced in the perceptions of *movement* that occur after one has been looking at a waterfall, or after one has been undergoing rotation (see pp. 180, 265).

This after-effect has also been observed in the perception of *direction.* If the head is kept turned to the right for a time and then brought back, a line directly in front of the observer

will be seen by him as at his left. The course of perceiving is, in this case, determined in part by kinesthetic stimulation from the muscles of the neck.

The perception of *curvature* shows similar changes. A curved line gradually grows less curved as it is being observed. If a straight line is then substituted, it will appear curved in the opposite direction. The following incident illustrates such change in curvature. In the upper story of an old university building the wooden stairs are hollowed out with use. After one has walked down these stairs, the remaining stairs, which have a new and even surface, are felt as being bumpy or curved upward. So compelling is the impression that one automatically adjusts one's manner of walking to the non-existing bumps and the walking becomes precarious.

These spatial changes point to an inherent instability of the central neural patterns involved in perceiving, the patterns altering in time. Such changes are useful, however, in that the negative after-effect serves to enhance slight differences between objects. Successive contrast, whether of colors or of spatial properties of objects, facilitates successive discrimination, just as simultaneous contrast is an aid to simultaneous discrimination.

Fluctuation. Another kind of instability results from the equivocal character of the stimulus. Fig. 98, for example, is seen at first glance as a disconnected assembly of patches of black. As one continues to look at the figure, however, it will suddenly alter and take on the appearance of a man smoking a cigar.

On page 231 there is a figure of one of the illusions of reversible perspective. The outline drawing is seen as a chair; but, instead of staying put, the chair reverses its orientation. Perceiving is instable, turning out now one product and now another. The stimulus itself is equivocal. Its characteristic spatial pattern compels the perception of a tridimensional object; but it is insufficient to determine a specific *orientation* of the object. If, however, the drawing is properly shaded, or if

a solid object is substituted for it, perceiving is stable; the chair stays put. The addition of new stimulus factors, such as light and shade, makes for greater stability of perceiving. Similarly, if a curved piece of wood were substituted for the curved line mentioned above, curvature adaptation would at least be considerably delayed.

FIG. 98. AN EQUIVOCAL STIMULUS.
Reproduced by permission of Winter Thomas Advertising Company, Ltd.

In Figs. 99, 100 and 101 one can easily observe a reversal of object and background. Here the mere difference between black and white seems to compel the perception of objects at *different distances,* but does not determine specifically what is to be seen as object in front and what is to be seen as background. Hence there is a shifting back and forth.

Reversal of object and background is likely to occur also when one is examining a plain map. One may at first take the

bodies of water for land masses. Inverted paintings and puzzle pictures also invite reversal of object and background. Other

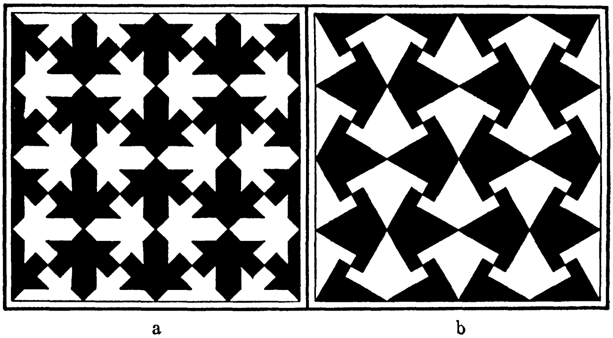


FIG. 99. REVERSAL OF OBJECT AND BACKGROUND.
From E. C. Sanford, A course in experimental psychology. Boston. D. C. Heath and Co., 1898, p. 254.

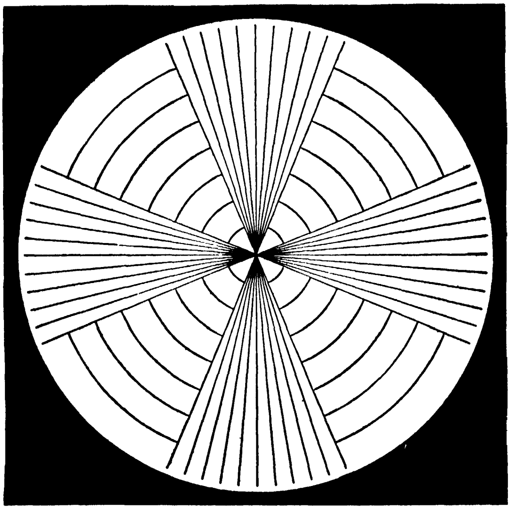


FIG. 100. REVERSAL OF OBJECT AND BACKGROUND.
From E. Rubin (12, Abb 5)

instances of reversal occur when one suddenly realizes that one has taken a line for a crack, a fleck of shadow for a stain, a fly on the window pane for a bird outside.

Retinal rivalry also furnishes interesting examples of fluctuation. If two diagrams differing in shape are presented in a stereoscope, it is difficult to get a composite view. Perceiving fluctuates from one of the single views to the other.

An example of auditory instability occurs when one is listening to a series of equal metronome beats. Even without any

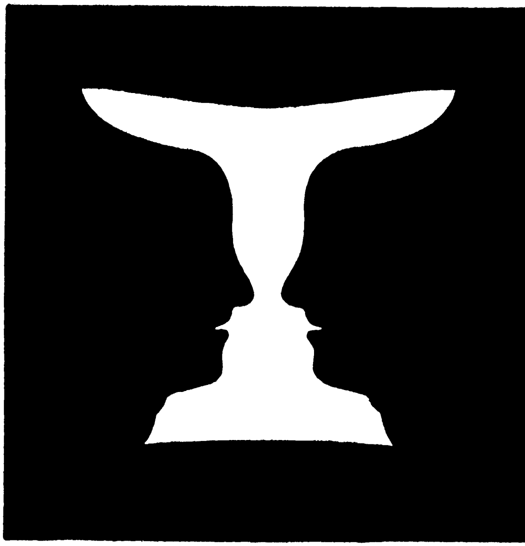


FIG. 101. REVERSAL OF OBJECT AND BACKGROUND.
From E. Rubin (12, Abb. 3).

rhythmical intention on the part of the listener, the beats are likely to fall into one rhythm after another.

THE TREND OF PERCEIVING

The final outcome in perceiving always depends upon the prevailing general trend of organic activity. This trend may work against perceiving. When one is preoccupied or when one is asleep, stimulation may be totally ineffective. In certain forms of hallucination, on the other hand, the trend encourages perceiving. One is likely to hear footsteps when expecting

a visitor, or the ringing of a telephone when one is awaiting a call.

In an experiment on color mixture, the observer was told that the blue sector would be gradually increased and the red sector diminished. Although the experimenter then proceeded to do exactly the opposite, the observer nevertheless saw the color as turning bluer. These instances, however, are but striking manifestations of the universal determining influence of organic trend.

Any change in the general trend of activity—whether self-initiated, as in the first two instances above, or initiated by suggestion or instructions from the outside—alters the character of perceiving. A child at play may use a wooden cube at one time as a missile, again as a building block and a third time as a locomotive. When we read "the quality of mercy is not strain'd," we do not think of a soup strainer. In these cases, the object or word is perceived *as the course of activity directs.*

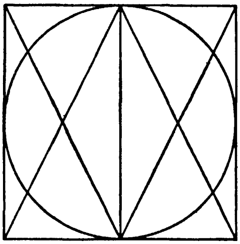


FIG. 102. REMODELING OF PERCEPTION.

Fig. 102 is likely to be seen at first glance as a square with an inscribed circle cut across by a few extra lines. Further inspection of the figure will reveal a diamond-shaped figure, 2 crosses, 2 rectangles, 2 parallelograms, 8 right triangles and 10 isosceles triangles. Perceiving is thus being *remodeled* by directed search or effort.

The following two experiments further demonstrate, and in a fairly striking manner, how a change in the trend of perceiving alters the outcome. In one experiment the observer faces a clock mechanism that is adjusted so that a bell rings when the hand of the clock has reached a certain position. If the observer *attends to the hand,* he will hear the bell at a *later* position of the hand, whereas if he is *set for the sound of the bell* he will hear it at an *earlier* position. Perceiving alters with the observer's *attitude.*

In the other experiment, letter groups differing in color and in arrangement (square, triangle, circle, etc.) are shown for ⅛ sec. The observer attempts at different times to note the total *number* of letters, their *colors,* their *arrangement* and their *individual character.* It turns out that the observer generally perceives only what he is set to perceive and fails to perceive anything else.

What happens in this experiment is usually characterized a *abstraction.* Only a few items are abstracted from the total situation. In this sense, perceiving is always abstractive. We never perceive at one glance all we can, even of a single object. When we *observe* an object, we are continually noting new features of it. Perceiving is altering and turning out new products. *Observation* is, in fact, a course of perceiving in which the output alters under the pressure of a trend that is directed to a definite end. We peer through the microscope in an effort to determine what lies beneath it. We inspect a fossil so as to classify it.

Many experiments in perception make use of this directive trend to alter the outcome of perceiving. In this way, clues may be obtained as to the conditions involved. For example, if, while perceiving wetness, one searches for *quality,* the experience may change from one of wetness to one of cold and pressure. And, conversely, when cold and pressure spots are stimulated simultaneously, one is likely to perceive wetness even though the stimulating objects are perfectly dry. This result shows that the stimulus is always equivocal; of itself the stimulus may lead perceiving to several different outcomes. It is the general trend that cooperates with the stimulus in determining the specific result.

In another experiment of the same sort dealing with the perception of pleasant and unpleasant objects, the search for quality revealed a diffuse bright pressure in the thoracic region of the body when the perceived object was pleasant, and a concentrated heavy dull pressure in the visceral region when the object was unpleasant. This result is important. It shows that the perception of pleasantness and unpleasantness involves a

reflex stimulation of organic receptors. In ordinary perceiving, however, this organic stimulation leads to the qualification of *objects* as pleasant or unpleasant rather than to qualities referred to the body.

In sleep the trend of perceiving takes many peculiar turns. One instructive instance is reported in an experiment on dreams. In this experiment, the subject, on being awakened, reported a vivid *visual* dream-perception of a horse. The experimenter, who stood by, had in fact heard the passing of a horse. In the subject's dream, however, the *auditory* quality was completely suppressed, although the stimulus was obviously effective.

The accompanying schema of the conditions of perceiving illustrates the role of the trend. The role of the stimulus in initiating and regulating the bodily processes involved in per-

CONDITIONS OF PERCEIVING

Stimulating energies	→	Transmission of energy to receptors	→	Transformation of energy in receptors	→	Neural impulses	→	Central repatterning of neural impulses	←→	Neural trend

ceiving is shared by the prevailing trend of neural activity. While normally initiated by stimulation, perceiving is constantly regulated from two sides: from the side of stimulus and from the side of organic or neural trend. The trend in its turn may, as is shown in the schema, be modified from the side of stimulation.

GROWTH OF PERCEIVING

In the last section we saw how perceiving changes with the momentary trend of organic activity. Earlier we found that, even when the trend is neutral or only favorable, perceiving may still be instable owing to the equivocal character of the stimulus and to the inherent instability of the neural-activity patterns directly involved in perceiving. The changes that represent the growth and development of perceiving remain to be considered.

Growth or development means more than a succession of changes. The changes must proceed in a definite direction, and they must be cumulative—the later changes implying the earlier. Development is thus a unitary temporal pattern of activity marked by definite stages.

Perceiving, as an organic function, is *continually* undergoing growth and development. On examining the outcome of this growth, we note that the way in which we perceive has been *remodeled* in a certain direction. We then say that we have *learned by experience.* When the child learns to avoid the flame that burned its fingers, its perceiving has altered. The flame comes to be perceived with a new significance.

This sort of remodeling takes place whenever we perceive an *event.* When witnessing a course of action—on or off the stage—when reading or when listening to conversation or music, our perceiving is undergoing a continual remodeling. When the crucial word in a sentence comes at the end, the meaning of the whole sentence may be suddenly transformed. The action, or the melody, takes on new significance as it progresses.

This aspect of perceiving, its development or temporal reorganization, can best be demonstrated by its limits. When a sentence is too long or involved, we finally fail to understand and we have to retrace our steps. There is a record of a man who, as a result of brain injury, failed to understand any but very short sentences. If the statement made to him was just too long, he would puzzle over the implication of the last phrase.

Development may lead to (1) an increase in the output of perceiving, (2) the use of new and more economical means, (3) the stabilization of perceiving. These three directions of development represent some of the different forms of remodeling or reorganization.

Range of perceiving. The range of perceiving has been measured in various ways. In one method, the method of testimony, a scene is presented on a screen for a moment. The

observer lists the items he saw, and his list is compared with the detailed list made out by the experimenter.

In another method a miscellany of letters, words, outline or block forms is exposed for less than 1/10 sec. through the window of a tachistoscope, an apparatus for rapid exposures. The results of such an experiment are given in Table XIV.

TABLE XIV

	I	II	III	IV
O_1....	11 3	7 9	4 3	3 3
O_2.......	6 2	6 9	3 2	3.0
O_3..........	9 2	5 9	3 9	3 3

All the figures represent the average number of items correctly reported in a series of observations. The different columns indicate the results obtained under different instructions. The 3 observers were required to report (I) *number of forms*, (II) *names of letters*, (III) *names of forms* and (IV) *number, name and color of forms*. The results represent the limits of perceiving only *when it is directed in a very specific manner*, for obviously we can take in with one look more objects than we can correctly number, name or describe. We can see thousands of people at once on a bathing beach.

These methods measure the output of perceiving in a small span of time. As perceiving is continued, or repeated, the amount perceived may increase. The practiced observer takes in more details at a glance than the unpracticed observer.

Economical modes of perceiving. We learn to *see* ice as cold and heavy, and materials as rough or soft. We come to perceive the weight of an object by the sound of its fall or by the groaning of its bearer, to identify a person by his voice, gait of the slope of his shoulders, to perceive a melody by the printed score. In all these instances a reduced stimulus becomes as effective as the more inclusive stimulus.

When one turns in the financial section of a newspaper to the chart that mimics, by an arrangement of lines, the ups and downs of the prices of stocks over a period of years,

one perceives at a glance a very long series of events. In this case spatial products of perceiving are put to temporal ends.

An economical change that has both advantages and disadvantages appears in the reduced form that is characteristic of our routine perceiving. We perceive very little of a familiar person, garment, room or scene. The various circumstances that are capable of breaking up this routine perceiving are often called the 'conditions of attention.' Sudden changes, the appearance of a novel object or of a familiar object in a strange setting are some of these conditions.

The stabilizing of perceiving. The perception of a new object or scene is at first often instable. There are many stimuli, some of which would result in the perception of one object, and some in that of another. Consequently there is much fluctuation, and perceiving is confused. On his first visit to the zoo, the small boy is distracted by all the various appeals made to his perceiving; the novice is helpless when set to using the microscope; one's first symphony is bewildering. In time, however, and with repeated perceiving, the conflicting stimuli cease to conflict; aspects of the objects which are not significant disappear, and fluctuation ceases. With each new perception there is also a gain in richness and completeness until ultimately the object has become familiar and stabilized.

Thus the boy at the zoo escapes his first confusion by perceiving only one of the many animals that surround him; all the others drop out and his perceiving is directed to that one. Then, as his eyes roam over the animal, as he walks around it and notes first one detail and then another, his perception is enriched. Finally his curiosity is satisfied and he passes on to another animal. The novice at the microscope has a similar experience. In a little while he is able to center his observation upon the field of the microscope, and details of the slide which have no significance for the object under study are ignored. As his perceiving is repeated new features are discovered and related to those previously observed.

The perceived object becomes more and more complete until at length the stage of stabilization is reached. With repeated perceiving, also, the symphony that at a first hearing is bewildering begins to take shape and the listener follows at will the melody carried by the violins, the echo of a phrase by the oboe, the announcement of a new motive by the brasses or the harmonic progression of the orchestra as a whole.

A similar example of this growth in the direction of stabilization is found in the repeated perceiving of a painting in an art gallery. At the first visit there is something of the same perplexity as in the other situations which have just been discussed. There are too many pictures demanding attention. At length some one canvas is singled out, and observation begins. The first perception may be only of a bit of color, an arresting figure or the subject of the painting as a whole. As the eyes wander over the picture new objects and special features of these objects—their design, their drawing, their color—are perceived in relation to the whole. Furthermore each new perception is modified by previous experience and by the trends of perceiving that are available to the observer. Everything he knows about painting—composition, values, drawing, technique—shapes and directs his tendencies to perceive certain features of the painting and their relation to each other. Ultimately his perception is stabilized and the picture has become an object vastly different from that of its first appearance; it is now an old friend to be visited and enjoyed with each subsequent visit to the gallery.

Practice in *discrimination* is largely a matter of such stabilization of perceiving. Suppose that pairs of tones differing slightly in pitch are presented repeatedly, and that the observer is asked to report 'same,' 'second is higher,' or 'second is lower.' If now the successive reports for a single pair of tones (the second of which is higher in pitch) are examined, one finds that correct discriminations appear early in the series of observations. The very first observation may lead to the correct report, 'second higher.' It is, therefore, not quite right

to say that the observer learns to discriminate. What does happen is that early in the series correct discriminations are few and far between, whereas later on they come more frequently. The observer becomes more consistent. At first, perceiving is instable; the two tones compared seem now equal and then unequal. After practice, the tones are more definitely and consistently the one thing or the other.

Growth is one of the means by which organic trends are established, and the normal trend of perceiving in turn favors

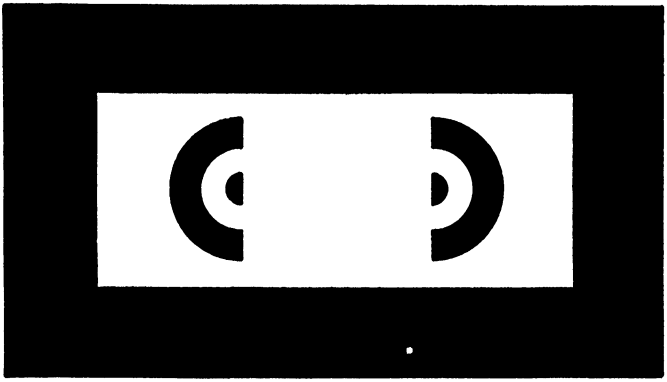


FIG. 103. A HIDDEN SQUARE.
From F. Schumann, *Zsch. f. Psychol.*, 1900, 23, p. 13.

the *stability* of perceiving. A lawn at a casual glance is seen as a uniform green although on careful inspection it shows many dull and drab patches. As a result of repeated perceiving we tend to see lawns as green; the perception has become stabilized. For this reason we may fail to obtain a reversal of object-background, or to discover the hidden face in a puzzle picture.

This resultant of growth explains in part the differences in perception among individuals when confronted with the same object. In Fig. 103, for example, one individual sees at once the white square in the middle of the figure while another fails to see it until it is pointed out. Similarly some persons are

quite unable to perceive the reversal in the staircase figure (Fig. 88); others cannot avoid the fluctuation.

Stabilizations of this kind are relative. Perceiving is generally stable because for most of our purposes the first perceptive impression suffices. But, whenever, as a result of curiosity or of instruction or of a need for more complete apprehension, an object is more carefully inspected, perceiving changes; stability then gives way to instability.

REFERENCES

1. BENTLEY, M. *The new field of psychology.* New York: D. Appleton-Century Co., 1934.
2. BRUNSWIK, E. Untersuchungen uber Wahrnehmungsgegenstande. *Arch f. d. ges. Psychol.*, 1933, 88, 377-418.
3. BRUNSWIK, E. *Wahrnehmung und Gegenstandswelt Grundlegung einer Psychologie vom Gegenstand her.* Leipzig and Vienna: Deuticke, 1934.
4. CANNON, W. B. *The wisdom of the body.* New York: W. W. Norton and Co., 1932.
5. GELB, A. Die 'Farbenkonstanz' der Sehdinge. *Handbuch der normalen und pathologischen Physiologie.* Berlin: Springer, 1929. XII (part 1), 594-678.
6. GELB, A., and GOLDSTEIN, K. *Psychologische Analysen hirnpathologischer Falle.* Leipzig: Barth, 1920.
7. KATZ, D. *The world of color.* London: Kegan Paul, 1935.
8. KATZ, D. *Der Aufbau der Tastwelt* Leipzig Barth, 1925.
9. KOFFKA, K. Perception: An introduction to the *Gestalt-theorie. Psychol. Bull.*, 1922, 19, 531-585.
10. KOFFKA, K. Some problems in space perception. *Psychologies of 1930.* Worcester, Mass: Clark University Press, 1930, Chap. 9.
11. MACLEOD, R. B. An experimental investigation of brightness constancy. *Arch. Psychol.*, 1932, 135, 1-102.
12. RUBIN, E. *Visuell wahrgenommene Figuren.* Copenhagen: Gyldendalske Boghandel, 1921.

CHAPTER 13

LEARNING

The psychology of learning includes two general division learning and retention. Each of these is a name given to a d' ferent aspect either of experience or of performance. Since, however, a majority of the experimental work classified under these headings has to do with the latter we shall, for the sake of simplicity, refer primarily to performance, but the student should remember that experience may also be involved. This chapter constitutes an introduction to representative problems in the psychology of learning and to the general results of experimentation upon them.

INTRODUCTORY CONCEPTS AND METHODS

Learning and retention. Human beings learn from birth to death, and the act of learning is so common as to go almost unremarked. When it is noticed at all, it is thought of in rough, unanalytic terms. The careful laboratory analysis of learning requires the use of materials, methods and concepts which may appear at first to be strange and divorced from the learning of everyday life. We should remember, however, that the learning with which this chapter deals is the same in kind as that which we are doing constantly, except that here it is dealt with in the terms of experimental analysis.

We measure learning and retention in terms of how much change in performance has taken place in a given unit of time or of effort. We give the name *learning* to an incremental change which results from practice and its attendant conditions. When a subject types more words per minute, when

This chapter was written by John A. McGeoch of Wesleyan University.

he types with greater accuracy (fewer errors) or when he repeats more of the words in a list as a result of practice, we say that he is learning. We give the name *retention* to any degree of measurable persistence of these increments through time, and the name *forgetting* to their failure to appear under measurement. From the standpoint of experimental measurement at any given time, retention and forgetting stand in perfect inverse relationship.

It should be emphasized that these terms refer to facts of performance, not to powers or faculties. Neither learning nor retention is an agent; neither does anything. Each is but a convenient label for observed or measured characteristics of performance. Each is to be understood in terms of the specific conditions under which the performance occurs and in terms of the operations by which it is measured. When, therefore, we employ the term 'learning' we mean a change in performance.

Learning and retention are not clearly separable characteristics of performance. The incremental changes which we call learning are at least in part retained from trial to trial. If they were not, the series of increments by which any given *criterion* of achievement, such for example as three perfect trials in succession, is reached would be impossible. Actually increments are retained or carried over from repetition to repetition and accumulate progressively. They are not, however, completely retained; they are in some degree forgotten, so that any given level of learned performance represents a balance in favor of retention.

There is no question but that retention and forgetting are present during learning; for experimental purposes, however, it is customary to make a distinction between learning and retention. When we are studying the course and conditions of learning, we disregard the fact that retention is also present and speak only of learning. When, on the other hand, we measure performance *at some time after the learning period has ended*, we call the results measurements of retention.

Experimental materials and methods. Anything learnable may, theoretically, serve as material for experimentation upon learning. Practically, however, except for learning under extra-laboratory conditions, experimental materials must satisfy certain criteria. They must be (1) relatively new to the subject; (2) divisible into equal units so that performance can be quantitatively measured; (3) learnable in the relativel short time at the disposal of experimental subjects; (4) of kind which can be brought within the bounds of laborat space and equipment; and (5) obtainable, preferably, in . number of samplings of equal difficulty or of known differences in difficulty. An understanding of these criteria will make clear the reasons for our use of certain learning materials and for our apparent neglect of others. There is good reason to believe that, although many of the things which human beings learn in practical life cannot be studied in the laboratory, the fundamental principles which are discovered in the laboratory are applicable to them. We shall, however, confine ourselves in this chapter to the experimental results and to illustrations in terms of laboratory materials.

Many materials meet these criteria and have long been in common use in the laboratory. One of the most widely used verbal materials is the nonsense syllable, which was first described in experiments upon learning by Hermann Ebbinghaus in 1885. It consists of a vowel between two consonants, as BAC or ZOK, and admirably satisfies the requirements. Lists of words, poems, pictures and similar materials also fulfil these criteria. On the side of motor skills, the stylus maze is a common instrument. It is a winding path from a starting point to a goal, with blind alleys branching off from it at a number of points. The subject is to find the true path without the aid of vision and to become able to go from starting point to goal without error. Typing, shorthand, telegraphy, mirror drawing, card sorting and ball tossing are other examples. Rational learning is studied by means of problems in logic, arithmetical reasoning, concept formation or the manipulation of spatial relations.

The material to be learned or the problem to be solved is presented to the subject under some definite instruction, and learning is measured either in terms of the amount acquired in a given time, or in terms of the number of repetitions required, the amount of time taken or the number of errors made in reaching a given level of performance. Any level of performance may be chosen as the one which the subject must attain: three errorless repetitions in succession is a common one. We call such a level a *criterion* or *criterial level of performance*, and when a subject has reached it we say that he has learned the material or mastered the problem.

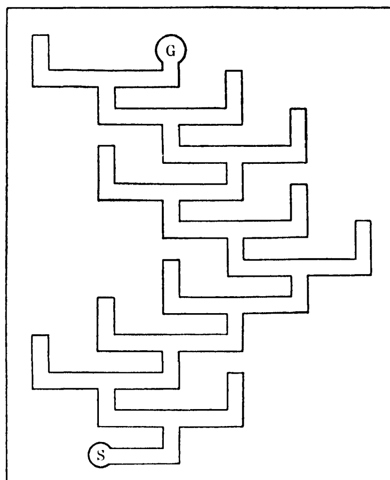


FIG. 104. A TYPICAL STYLUS MAZE PATTERN. From C. J. Warden, *J. Exper. Psychol.*, 1924, 7, p. 101.

A classification of things learned. It will have been evident from the description of experimental materials that they may be conventionally grouped in three classes, *viz.*, verbal materials, acts of skill and rational problems. The three are far from being clearly distinct groups, however. Many laboratory materials belong in part to two or sometimes to all three of the groups. The classification, nevertheless, offers convenient points of reference both for experimental materials and for the activities involved in learning them.

It is a general characteristic of experimental work with verbal materials that they are presented to the subject in the form in which he is to learn them. Learning, in this case, is a *fixation* of the materials by the subject, and fixation means be-

coming able to repeat or use them without the copy. In the case of acts of skill, such as the running of a stylus maze, however, the subject is commonly presented with the end-result to be achieved and with certain general information concerning the means to be employed, but he must first find for himself specifically how to reach the goal. This includes finding out what movements will attain it and how to make them. In the discovery of the correct movements many wrong or useless ones will also be made, and attainment of a criterion of mastery requires that the wrong moves be eliminated and the right ones fixated. The acquisition of skill, thus, adds *discovery* and *elimination* to the fixation which is the dominant feature in acquiring verbal materials.

The learning of rational problems has these same characteristics but differs from the learning of acts of skill in the devices used in the discovery of the solution. The solution of a problem in skill is achieved by a series of overt movements, whereas the solution of a rational problem is reached by means of sequences of ideas. This distinction is far from being absolute, but it is useful. Problems in skill involve ideas as well as movements, and rational problems involve overt movements as well as ideas; but the one is more characteristically motor and the other more characteristically ideational. In both there is the variable pattern of right and wrong acts with the final fixation of the right ones and elimination of the wrong ones.

Distinctions between the activities required by the three groups of learning materials are relative. Things learned may be distributed along a continuous abscissa with respect to the amounts of discovery and elimination which they require and with respect to the extent of overt movement and of ideation involved. The distinctions made in the succeeding paragraphs offer, however, an introduction to some of the major descriptive aspects of learning.

The common use of these distinctions has led to the giving of separate names to the time or repetitions devoted to learning the different materials. The repetition of verbal material is called *memorizing*; the repetition of an act of skill is called

practice, and the repetition of attempts to solve a rational problem is termed *reasoning.* We shall use the term *practice* as the generic term and shall employ the two others only occasionally. The three terms are applied when there exists a reasonable expectation of further increments in performance as a result of the repetition. When the repetitions are made under conditions which do not offer a reasonable expectation of, or opportunity for, improvement, they are designated as *work.*

Trial and error. The series of right and wrong acts by which the goal is discovered in the learning of both skill and rational problems is known as *trial-and-error behavior*, and the performance of such a series is termed a *trial-and-error method.* Everyone employs this method of attack to some degree. The student can grasp clearly what it involves by recalling his own behavior in trying for the first time to solve a puzzle, to open a box with a hidden catch or to find his way to a given address in a strange city when only the general location is known.

A considerable amount of misunderstanding has grown up around this concept, and it requires, therefore, some elaboration. (1) It should be noted that trial-and-error behavior is not purely random behavior. The subject proceeds under some instruction, which is either given to him by the experimenter or imposed by himself, and this instruction limits and to some extent controls his behavior. A further control is provided by the previous training and structural characteristics of the learner which bring it about that certain responses will be elicited by the learning situation and that others will not. There is always some internal guidance of behavior, and the only purely random or chance aspect of trial and error lies in the fact that neither subject nor experimenter can predict before the practice begins the time at which the goal will be reached or the number of errors which will be made in reaching it. *The behavior itself is variable but not at random.* As practice goes on, variability becomes less, until it is finally re-

duced to small changes in the way in which the right response is made. (2) Trial-and-error is a descriptive term, not an explanatory one. It describes the behavior exhibited by the learner in attacking a problem, but it does not state the conditions necessary for final mastery.

FORM OF LEARNING FUNCTIONS

Representative curves. The first question to arise, when we turn from a description of things learned to the conditions of their being learned, is the influence of successive amounts of practice. Few acts or materials can be acquired at a single trial or single brief moment of time. Most of them require several repetitions and considerable time, and with succeeding repetitions varying increments of performance appear. We wish to know the relation between successive units of practice and the performance which practice brings. This relation can be seen most clearly if we plot performance against repetitions and connect the plotted points. The resulting graph is called a *curve of learning*. When a relatively small number of trials has been required, every one may be represented in the graph. When a large number is necessary, only every fifth or every tenth record may be plotted; the resulting graph is a curve of sample performance, not of total performance.

Learning curves may show a general rise or a general fall depending upon the way performance is measured. If measurement is in terms of time or of errors, the curve will fall; if in terms of amount accomplished, as number of words learned, it will rise. The functions for individual subjects will show in either case a considerable fluctuation. This is illustrated by the curve in Fig. 105 which represents the number of nonsense syllables recalled at each successive presentation of a 10-syllable list. These fluctuations are a result of variable factors in the stimulating conditions and in the subject's response to them and are virtually inevitable features of performance during learning.

Sometimes fluctuations may continue for a relatively long

time without showing any trend toward further increments. This region of no definite change in performance is called a *plateau* (Fig. 106) and is of considerable interest because of its departure from the general trend of the functions and, as well, because of the discouragement which it causes the learner.

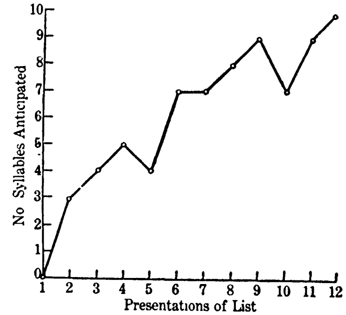


FIG. 105. A MEMORIZATION CURVE FOR ONE SUBJECT.

Plateaus are not to be accounted for by any single condition. They may be produced by a decrement in the subject's motivation and in the vigor with which he practices, by interference between the different parts of the reactions which are being learned, by the fixation of errors or by any condition which will for a time retard performance. It is certain that they are not necessary features of the relation between practice and performance.

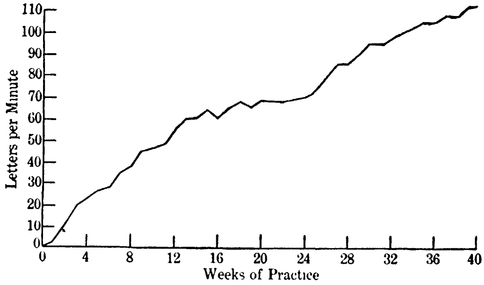


FIG. 106. A PERFORMANCE CURVE FOR ONE SUBJECT IN LEARNING TO RECEIVE IN THE TELEGRAPHIC LANGUAGE.

"This is a curve of sample performance. The region of very little or no progress toward the middle of the curve is a plateau." From W. L. Bryan and N. Harter, *Psychol. Rev.*, 1897, 4, p. 49.

The fluctuations which typically occur in the curves for individual subjects average out to give a smooth composite curve when the functions of a large number of subjects are

combined. Both irregular individual curves and the smoother composite ones usually approximate one of the three characteristic forms of Fig. 107. There are, however, as many specific curves as there are conditions under which learning is done, and the three given are only general modes with respect to which the variety of empirical curves may be classified and described. There is, therefore, no curve which may be called *the* curve of learning, none which represents learning as a whole or in general. Each curve is a geometrical picture of the relation between some specific measure of performance and some specific measure of practice. Each is relative to the conditions under which it occurs.

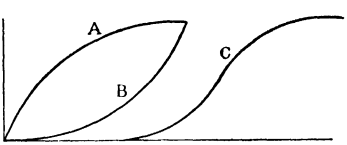


FIG. 107. REPRESENTATIVE LEARNING CURVES.

Curves *A* and *B* are from Carr (1, 218).

The chief problem of description in the present case is that of the conditions under which a given curve-form appears. Negatively accelerated curves of Type A show that the increments of performance are relatively large during the early stages of practice and relatively small during the later ones. Curves of this form are most often obtained (1) when motivation is high at first and decreases as practice continues; (2) when interference-effects, within the materials being learned, increase as practice goes on; (3) when the subject has had previous practice, the effects of which transfer to the early repetitions of the material being learned or to the learning method employed; and (4) when measurement is in terms of errors or time rather than of positive increments of performance.

Positively accelerated curves of Type B represent increments which are relatively small during the early part of practice but which increase in magnitude with continued practice. We seldom find curves which are positively accelerated throughout (they appear only when the limit is an arbitrarily selected criterion). More commonly we find an initial positive acceleration which may continue for some time but which

is followed by negative acceleration. In this case the total curve is an S-shaped curve of the kind shown as Type C. Curves of Type B and C most often appear (1) when the subject has had very little prior practice, the effects of which transfer to the initial repetitions of activities concerned; (2) when the subjects are relatively young, either chronologically or mentally; (3) when the acts learned are, for any reason other than those already mentioned, relatively difficult for the learner; and (4) when scoring is in terms of positive attainment rather than in terms of time or errors. Occasionally we find curves which are straight lines (zero acceleration). They are usually sections of a long curve of one of the other types and are produced by combinations of the conditions which yield these types.

Physiological limits. A consideration of learning curves leads directly to the question whether or how rapidly learners reach a physiological limit, *i.e.,* the level of performance beyond which, by reason of the physical limitations of their organisms, they cannot go. Under ordinary laboratory conditions subjects seldom approach such a limit of performance, and one must rely for an answer to the question upon somewhat fragmentary evidence. It has been found that years of practice at such skills as telegraphy or typesetting do not commonly bring a man to his maximal performance. The introduction of a special incentive may increase the performance of long-practiced workers by remarkable amounts. The fact that in athletics and in other skills records are repeatedly broken under standard conditions is best interpreted as an indication of the practical remoteness of a physiological limit. It has, likewise, been found repeatedly in the laboratory that, after a subject has reached a relatively high level of performance, increased motivation or better methods will produce further substantial increments. The conclusion is probable that in normal persons a physiological limit is not reached by ordinary amounts of practice and, indeed, may not be reached even by prolonged practice under favorable conditions.

LEARNING AS A FUNCTION OF THE CHARACTERISTICS AND BEHAVIOR OF THE LEARNER

The problem. The speed with which a given level of performance is reached offers a problem entirely distinct from that of the correlation between successive increments of performance and successive units of practice. The criterial level might be attained in 5 trials or in 20, and the form of the curve might be the same in either case. The general problem involved in studying speed of learning apart from the form of the curve is the problem of the conditions which determine speed and becomes, when stated in experimental terms, the measurement of either the amount of increment in a given number of practice units or the number of practice units required to attain a given level of performance. The conditions which determine a given rate may be roughly classified into those which vary with or are a function of the learner and those which vary with or are a function of the material learned and the conditions of its presentation. We shall deal in this section with the conditions in the first group and shall summarize briefly in each case the more important conclusions of experimental work.

Chronological age. Under normal conditions, rate of learning increases as chronological age increases from the earliest age at which adequate measurements can be made to some point in the 'teens or early twenties. The form of the curve obtained by plotting performance against age varies with a number of conditions, as does the point at which further increases in age cease to be accompanied by increases in rate; but the fact of the increase in rate from an early age to some later one remains unaffected by a very wide range of conditions. The curve for card sorting shown in Fig. 108 is an example. The data plotted are obtained, as is commonly done, by measuring different groups of subjects at each age. We assume that the resulting trend approaches what would be found by successive yearly measurements of the same subjects.

The question of the form of the curve beyond early maturity has received extensive experimental attack only within recent years. Increases with age in the learning of most of the activities studied have ceased to appear by age 20 or soon thereafter. The evidence which is now available indicates a gradual decline from the early twenties onward to old age. The curve given in Fig. 109 for code learning is representative of this decline.

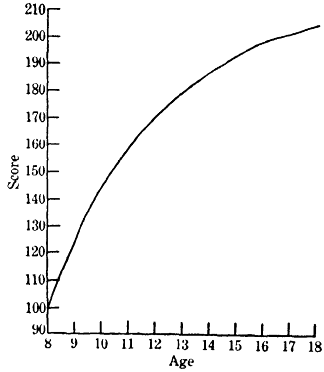


FIG. 108 A SMOOTHED CURVE SHOWING CHANGE WITH AGE IN PERFORMANCE DURING PRACTICE AT CARD SORTING.

From W H. Pyle, *Nature and development of learning capacity*, Baltimore. Warwick and York, 1925, p. 19.

Learning in most activities probably decreases gradually with age from early maturity onward, but this does not mean that the older person cannot learn, nor need it be interpreted to mean that he is seriously handicapped. It means rather that he learns less readily than he did at the peak of his performance. In code learning, for example, the person of 60 learned as rapidly as the child of 11, and the 11-year-old is not usually considered an ineffective learner. In a few activities and under a few conditions, moreover, it seems that the age curve shows

FIG. 109. THE TREND OF MEAN PERFORMANCE IN DIGIT-SYMBOL SUBSTITUTION WITH AGE.

From R. R. Willoughby, *J. Educ. Psychol.*, 1929, 20, p 678.

either no later decrement or shows it very late in life. The acquisition and manipulation of ideas, particularly if they are related to fields of knowledge with which one has been actively engaged, offer a case in point.

Intelligence. One might well expect, on the ground of common observation, a high correlation between intelligence and learning. When, however, intelligence-test scores and measurements of performance upon specific learning tasks are correlated, the coefficients are seldom high. The relation is, nevertheless, almost universally positive. When, moreover, measures of performance upon a number of learning tasks are combined and then correlated with intelligence-test score, the coefficient is a substantial one ($r = 0.60$, for example).

Evidence is accumulating that the magnitude of the relation between intelligence and learning is a function of the task learned and that the relation becomes increasingly high as we pass from simple tasks to complex ones. The highest coefficients yet obtained have been with rational problems where the coefficients may reach the 0.80's. It has been shown, further, that the influence of practice upon the differences in performance between groups of different intelligence is a function of the task. The learning curves for card sorting converge with continued practice, those for code substitution remain parallel, while those for the learning of abstract relations diverge.

We may conclude that speed of learning and test-intelligence are, over a wide range of things learned, positively related, that the magnitude of the relation is a function of the learning task and is much larger when records for several tasks are combined than when each is correlated individually.

Motivation. We shall use the term 'motive' to designate *any condition of the organism which points it toward the practice of a given task and which defines the satisfactory completion of that task.* Conditions such as hunger, thirst, desire for social recognition and avoidance of pain are examples of motives. It is obvious that a subject will practice only under the influence of some motivating condition and that the

'right' acts, *i.e.,* the acts he attempts to learn, are distinguished from the 'wrong' ones, or are defined in terms of their relation to the motivating conditions.

In experiments with human subjects the motivating conditions are complex. A college student may practice a maze habit under the influence of a motive to please the experimenter, to make a record which he himself will consider good, to excel over other subjects or to find out what maze learning is like. Any one or all of these may be operating in a single subject, and we may add to them, in a given experiment, the motive to escape an electric shock administered whenever a blind alley is entered or to receive some specific reward when the maze is learned.

The form which the problem has commonly taken in experimentation has been a comparison of learning under whatever motivation is present when a subject is merely asked by an experimenter to learn a given task, with learning under this condition plus some special motive such as reward or punishment. The introduction of such special motives almost uniformly accelerates learning, and the extent of the acceleration is sometimes large. In one experiment, for example, an electric shock of 27 volts, administered at the end of each blind alley of a stylus maze, decreased the number of trials required to reach the criterion by 50 per cent and decreased the various kinds of errors by 30 to 60 per cent. Competition and rivalry, praise and blame, and similar special conditions designed to arouse special motives have, likewise, exerted a definite facilitating effect. There is no doubt that over a wide range of special motivating conditions this result may be expected.

The general problem of motivation has also been studied, especially with verbal and observational materials, under the rubric of the influence of intention to learn. A subject may observe or repeat a material passively or he may assume toward it an active attitude and attempt to learn it. The experimental results show that mere passive repetition is either without influence or is relatively ineffective and that active noting and

reacting to the material are necessary for learning. At least a part of the effectiveness of motivation is referable to the intent to learn which it produces. (For a further description of motives see pp. 456-459.)

Recitation during learning. A subject may practice by repeating the act or material as it comes or he may attempt to recite or rehearse it during the learning with prompting as necessary. Thus, during the running of a maze he may verbally rehearse the turns which he has just made while he is at the same time going forward through other sections, or during the memorizing of a list of words he may try to repeat the list without the copy. Table XV presents data upon the influence of varying amounts of recitation. Lists of nonsense syllables and short biographies were practiced either by repeated reading until learned or by reading followed by recitations with prompting when necessary. The table gives the percentages recalled immediately at the close of the learning period.

TABLE XV

THE INFLUENCE OF DIFFERENT AMOUNTS OF RECITATION UPON LEARNING

(From Gates, *Arch. of Psychol.*, 1917, 4, pp. 36, 41)

Per cent of total time spent		Material learned	
In reading	In recitation	Syllables	Biographies
100	0	65 4	87 8
80	20	92 2	94 6
60	40	99 7	105 0
40	60	105.5	105 5
20	80	137 3	106 8

(The figures are for subjects in Grade VIII and have been obtained by computing the percentage which the amount learned by each method is of the average of all methods)

These results show that direct repetition plus recitation yields larger increments of performance than time spent only in direct repetition, and that the increments vary directly in magnitude with the proportion of the total learning time spent

in recitation. The advantage of recitation is greater with the nonsense syllables than with the biographies, probably because a considerable amount of organization of meanings is progressively going on during the repetition of meaningful material even when there is no overt attempt at recitation. It has been found by other investigators that more favorable results are obtained when readings and recitations are interspersed than when they are grouped together.

The superiority of reading plus recitation over reading alone results from several independent conditions. (1) The recitation arouses a more active attitude on the part of the subject, and this is known to facilitate learning. (2) During recitation the subject is practicing the recall of the material in the way in which he is to use it on the test. (3) The recitation yields progressive information about both errors and right responses, thereby permitting the correction of the errors through prompting from the copy and providing increased motivation. We must note that recitation, as it is usually studied experimentally, involves prompting also, and that the presence of this factor must be taken into account in interpreting the results.

Guidance. The usual procedure in the learning of acts of skill and of rational problems is to allow the subject to discover and fixate the solution by his own methods. He must therefore proceed by trial-and-error. It is possible, however, in a great many learning tasks for the experimenter to give the subject aid during discovery and fixation. This aid is designated as *guidance* or *tuition*, and is distinguished from the guidance provided by such factors as the subject's self-instructions and transfer from his past training in that it is external, that is to say, it is more directly imposed by the experimenter, whereas the latter is internal and a function of the learner. A series of studies of external guidance has been made, using both stylus mazes and rational problems. In the case of the maze, the subject may, for example, be guided a certain number of times through the true path, may be prevented from making errors by blocks inserted in the blind

alleys, or may be given verbal information. When rational problems are used, the fundamental principles required for solution may be stated or the subject may be given information designed to aid him in finding the principle.

In a majority of the maze experiments some guidance has proved to be more effective than an equivalent amount of unguided practice, although occasionally guidance has been either ineffective or detrimental. The greatest positive effect is obtained when small amounts of tuition are given early in practice. Large amounts and amounts introduced late in practice have usually only a very small positive influence or are actually harmful. In experiments upon rational learning, information leading to the principle involved in the solution, if given early in practice, may act to hasten learning.

Individual cues and modes of attack. We have been dealing thus far with objective performance and its objective conditions. Experiential factors should also be considered. The reports of subjects upon the experiential accompaniments of increments in performance do not provide us with a point-for-point experiential parallel to the course of learning. A wisp of imagery, a faint imaginal schema, intermittent flashes of subvocal speech, a fluctuating series of kinesthetic strains variously localized are, for example, what the subjects report. So slightly correlated are these phenomena with those of performance that their importance is difficult to assess. The surprising and significant thing is that the experiential accompaniments of learning are so sketchy and unsystematic.

One form of investigation has yielded a definite relation between measured performance and subjective report, but here the reports are type reports rather than specific ones. Subjects who have learned a maze are asked to describe the modes of attack or the experiential means which they used in learning. Their reports can be grouped in three basic divisions: (1) verbal methods, in which the turns and other moves are carried in verbal form, so that the subject guides himself through the maze by saying, "first turn to the right, then straight

ahead," etc.; (2) motor methods, in which the cues employed are dominantly kinesthetic and the subject "follows the lead of his hand" without organizing his movements under the pattern of representative devices; and (3) visual methods, in which attempts are made to construct visual images of the pattern. The frequencies with which the three methods are used decrease in the order just given, with the visual methods appearing infrequently. The learning records of the subjects using the different methods show a clear superiority for the verbal method over the two others and, usually, a superiority of visual over motor. Table XVI gives a sample set of results for one section of a high-relief finger maze which the subject learns to trace with one finger. Results of a similar tendency have been found with stylus mazes, where it has been observed, also, that the more intelligent subjects are the more apt to adopt the verbal method.

The pronounced superiority of the verbal method over the motor in the learning of a motor problem is very significant. It shows that the motor problem as actually learned by a large percentage of the subjects is learned as a pattern with both representative and motor constituents, that so-called 'motor' learning is not, therefore, of necessity completely motor, that the trials and errors may be in part ideationally controlled, and that it proceeds much more rapidly when ideational factors are employed. These facts support the generalization made earlier in this chapter that there is no clean-cut division between different kinds of materials with respect to the activities

TABLE XVI

PERFORMANCE SCORES MADE BY SUBJECTS USING DIFFERENT MODES OF ATTACK

(Data from Husband, *J. Genet. Psychol.*, 1931, 39, 258–277)

Mode of Attack	Trials	Score (errors)	Time (sec.)
Verbal	10 1	20	358
Visual	15.0	29	505
Motor	25.8	23	802

which they require. Clear indication is also given of the importance and pervasiveness of representative factors in learning.

Individual differences. Wide differences are found between the learning records of different individuals when all are practicing under the same external conditions. These differences still appear when the subjects are of the same age and sex, in the same class in college, and of equivalent prior training upon the activity practiced. In one experiment in which these conditions were fulfilled, the fastest learner required 8 trials to learn a list of 8 nonsense syllables and the slowest required 37, or more than 4 times as many. Likewise, the fastest learner mastered a maze of considerable difficulty in 19 trials, whereas the slowest took 78. Ranges still wider than these are often obtained. When the records of a large group of subjects are plotted, the resulting distribution curve usually approximates the normal probability curve (pp. 41f.).

THE CHARACTERISTICS OF THE ACTS OR MATERIALS LEARNED AND OF THEIR MODE OF PRESENTATION

The problem. We shall be concerned in this section with the influence upon rate of learning exercised by conditions which are primarily functions of the things learned and of the way in which they are presented to the subject. Learner and thing learned are not, of course, as sharply divided from each other as this classification makes them seem to be. What is learned is not a purely external something to be absorbed by the learner, a something which is the same for everyone. It is rather something perceived by the subject, and it will vary as the subject varies. When, however, we systematically vary the characteristics of the material, while maintaining a chance distribution of subjects, we are able to treat the material as the major variable and to study its influence.

The meaningfulness of the material. If we adopt for the moment a common-sense view of meaning, it is possible to rank a large number of verbal materials from low to high with respect to their meaningfulness. On such a scale, nonsense sylla-

bles are placed well toward the lowest end, single words have a higher ranking, poetry or prose a still higher one. An almost perfect correlation is found between meaningfulness and ease of learning, so that it may be said that over a wide range of materials rate of learning is a direct function of the meaningfulness of the material. This does not mean, of course, that the meaning does anything, nor does it signify that the meaning in question is divorced from past learning. It, in turn, is a function of the subject's prior learning of related things, so that in one sense the generalization just made could be stated to read that rate of learning is a direct function of the amount of transfer which takes place from the learner's already existent knowledge or organization to the particular performance at hand. This factor has a much greater influence than do such physical features of the material as the size of the letters and the color of printing or background. These latter conditions may sometimes affect rate of learning, but the magnitude of their effect is almost negligible compared to that of meaning.

The meaning factor may also be approached from the side of the activity of the learner. A subject may study a list of words or a paragraph of prose in verbatim fashion, attempting to connect each word with the succeeding one, or he may try to weave the words into a meaningful pattern in which the interrelations either aid in fixing the words or entirely take precedence over them. The two methods are usually referred to as the *verbatim* and the *logical* methods of memorizing, and in the majority of cases the latter yields the more rapid learning.

Amount of material. The question of the relation between the time required for learning and the length or amount of the material learned has received considerable study. The common result has been that material becomes disproportionately more difficult to master as it increases in length. One investigator found, for example, that he could learn a list of 12 nonsense syllables in 1.5 min., a list of 24 in 5 min. and a

list of 72 in 25 min. If time had been proportional to length, the corresponding figures would have been 1.5, 3.0 and 9.0 min. This generalization has been found to hold for a large number of verbal materials and at least partially for maze patterns.

The disproportionate increase in difficulty with increasing length may be accounted for by the operation of several factors which vary with length. As length increases, the opportunities for interference between different parts of the material and for consequent forgetting increase. Some parts of a material are learned early in practice but are unnecessarily repeated during all later trials. The shorter the material, the less will there be of such repetition. The attitude of the subject toward the practice is also more favorable when the amount of work before him is relatively small.

Distributed vs. massed practice. The trials required to bring performance up to a given level may be taken in immediate succession, in which case the practice is said to be *massed*, or intervals of no practice may be inserted at any desired points to give a *distribution of practice*. These two conditions may be regarded as methods of learning or as modes of presentation. We shall take the second alternative and consider massed practice as involving continuous presentation of the material, and distribution as involving intermittent presentation.

The two chief variables in distributing practice are (1) the number of trials (or amount of time) per practice period and (2) the length of the interval of rest, *i.e.*, of no practice, between practice periods. In experiments upon the influence of the first variable a constant rest interval is used, and various numbers of trials are taken before each introduction of the interval. A comparison of 1, 3, 5 and 7 trials before each rest interval of 24 hr. would be one example. In experiments upon the second variable, on the other hand, the number of trials is kept constant and varying lengths of rest interval are compared. Thus, intervals of 1, 12, 24 and 48 hr. might be inserted after each practice period of 3 trials. Other forms of combination of these two variables are possible, such as an increasing series

of intervals and a concomitantly decreasing series of lengths of practice, or the insertion of only a single interval at varying points during practice.

The conclusion is well established that, over a wide range of conditions, some form of positive distribution is a more favorable condition of learning than is zero distribution, or massed practice. The curve shown in Fig. 110 is one illustration of this fact in the case of an interval as short as 1 min. The differences between performance under massed and distributed practice are commonly greater than are those between any two modes of distribution thus far considered. These statements refer, of course, to comparisons of amounts of time actually spent in practice. The total time covered by the experiment will almost always be greater under distribution because of the addition of the rest periods.

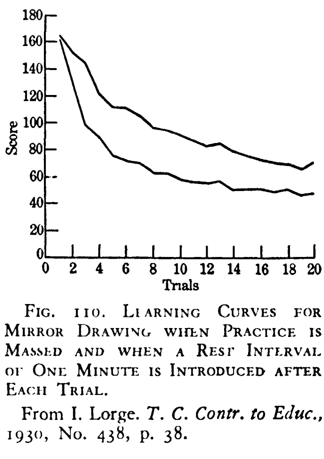


FIG. 110. LEARNING CURVES FOR MIRROR DRAWING WHEN PRACTICE IS MASSED AND WHEN A REST INTERVAL OF ONE MINUTE IS INTRODUCED AFTER EACH TRIAL.

From I. Lorge. *T. C. Contr. to Educ.,* 1930, No. 438, p. 38.

The relative effects of various numbers of trials and various lengths of rest intervals are functions of a number of conditions, and generalizations concerning them can be understood only in terms of these conditions. A few of the more general results may be stated without, however, a detailed analysis of the related conditions. (1) The influence of the number of trials per period is a function of the length of the rest interval between periods, but when an advantageous interval is used the fewer the trials per period the more rapid is the learning likely to be. (2) Relatively short intervals are more often advantageous; relatively long ones are more often detrimental. Very long ones, particularly those which are longer than a few days, are almost always detrimental. The longer the practice period, the longer may the effective interval be. (3) For many

activities, distribution is most effective during the early practice periods; and for some, a progressive decrease in practice period with increasing length of rest interval as practice continues is more effective than an unchanging series.

The primary conditions which determine the facts of distributed practice offer an interesting problem, particularly in view of the fact that an interval of no practice is commonly accompanied by forgetting. It has been demonstrated that rehearsal in the rest intervals of distributed practice does not account for the results. No single factor which will account for them has been found, and it is probable that they are determined by a group of conditions, not all of which need to be operating in every case. Some of the more important of these, all of which deserve further experimental analysis, may be mentioned. (1) The rest intervals may provide an opportunity for the dropping out or forgetting of wrong associations, especially when acts of skill are being learned. These associations are commonly less strongly formed than the right ones and hence are forgotten more rapidly. (2) Jost has formulated the hypothesis, which has been supported in several specific instances, that, if two associations are of equal strength but of different age, a new repetition strengthens the older association more than the younger. In distributed practice, all associations will have a greater age at all practice periods after the first than at the corresponding repetitions under massed practice and hence will be relatively more strengthened by succeeding repetitions. (3) In some cases, also, motivation may be higher and the subject's set more favorable to learning when the practice periods are relatively shorter under distributed than under massed practice. (4) It has recently been found that associative processes exhibit a phenomenon of *refractory phase* in that an associative response, once made, tends not to be repeated at once. There seems to be a barrier against immediate repetition. This fact favors distribution. (5) So, also, does the fact that the names of tasks, and perhaps the tasks themselves, which have been interrupted before completion, are better recalled than names and tasks which have been completed.

Other conditions. We have been examining representative samples of the conditions which affect speed of learning. There are many others, some of which have a positive influence, some of which are negative. Rhythm, for example, aids learning importantly. Artificially induced muscular tension during learning often has a positive effect. Whether practice of a material as a whole or in parts which are later connected, as in learning poetry a stanza at a time, is the more effective, is a function of too many conditions to permit a single generalization. The sense department to which the material is presented has little effect upon rate of learning. The details of these and of other conditions must remain for subequent discussions of learning.

INTRA-SERIAL RELATIONS

Serial position. Many of the things learned under laboratory conditions are arranged in series of more or less equivalent units. However, it is a well-established fact that items in some parts of the series are learned more rapidly than items in other parts. The data of Table XVII represent the usual relation between serial position and learning. Items in the early and late positions are learned most rapidly, while those in the intermediate positions are learned more slowly. The least advantageous positions are those just past the middle of the list. The alleys of a maze show the same relation, although with somewhat less regularity of change from position to position. The superiority of the early and late positions is known as the *primacy-finality effect.*

TABLE XVII

PERCENTAGE OF CORRECT ANTICIPATIONS AT EACH SERIAL POSITION DURING 10 PRESENTATIONS OF A LIST OF 10 ADJECTIVES

Position in series	1	2	3	4	5	6	7	8	9	10
Percentage correct anticipations	71	63	58	53	47	43	47	45	51	68

It is probable that the primacy-finality effect is largely a function of the greater freedom of the first and last parts of the list from interference of a kind which we shall

presently discuss under the name of associative and retroactive inhibition.

Remote forward and backward association. The direction of practice upon serial activities is forward. It is, however, a question of great importance whether the learning is entirely in a forward direction and whether it is a matter of connecting or associating adjacent terms only, or whether backward associations and associations between serially remote or non-adjacent items are also formed. The experimental work on the problem has shown clearly that associations are formed in the backward direction as well as in the forward and that a considerable number of remote associations are formed in both directions. The largest number of all is between adjacent items, however, and the frequency of the remote connections declines in most cases with increasing degree of remoteness. These facts mean that the members of a series are complexly interconnected to form a much more highly organized whole than serial connection alone could give.

The possibility at once presents itself that these remote associations were formed by direct practice in the sense that the subject reacts to spatially remote items in direct temporal succession. This could be done by recalling a remote item while any given item is present in perception. The best evidence available, however, indicates that this explanation will not account for all cases and that there are conditions where it will account for no more than a few. Remote associations may, apparently, be formed indirectly and across the time-gap filled by the subject's reaction to intervening items.

INTER-SERIAL RELATIONS

The influence of practice upon successive samples of material of the same class. Practice upon one sample, such as one maze or one list of words, of a given material may result not only in the learning of that sample but also in a facilitation of the learning of other samples of that specific material. Practice effects of this kind, which are not specific to

the sample learned, are one form of what is called *transfer of training*. They occur within most materials, probably within all, and they are often large. The data of Table XVIII illustrate such transfer. They are for the learning of a problem in which the subject has first to discover the correct responses and then to fixate them, using the information gained at successive responses, to avoid later errors by a rational procedure. In a certain maze experiment, likewise, the number of trials decreased from 17 to 6 after practice upon two other mazes, time decreased from 676 sec. to 127, and errors from 194 to 35. With continued practice at successive samples, the increments in rate of learning continue to appear after a relatively larger number of samples but at a negatively accelerated rate.

TABLE XVIII

THE INFLUENCE OF PRACTICE AT ONE AND AT TWO PRIOR RATIONAL LEARNING PROBLEMS

(From McGeoch and Oberschelp, *J. Gener. Psychol.*, 1930, 4, 164)

Number of Prior Problems	Pairs Presented	Time (min.)	Errors: Logical	Errors: Perseverative	Errors: Unclassified
0	112.50	32.32	104.62	24.25	121.75
1	75.00	18.33	44.62	14.75	74.25
2	58.50	10.66	22.75	2.00	63.37

(The Peterson Rational Learning Problem consists of a series of pairs of letters and numbers such as $\begin{matrix} A & B & C & D & E & F \\ 4 & 1 & 5 & 3 & 6 & 2 \end{matrix}$ The experimenter calls out the first letter and the subject is to guess which number between 1 and 6 goes with that letter, and so on to the end. A number once correctly given cannot logically be used again, nor can the same wrong guess be repeated. The subject must learn to make the correct pairings without error.)

Bilateral transfer. Bilateral transfer, or cross-education, a second form of transfer of training, refers to a facilitation of the learning of responses on one side of the body by the learning of responses made on the other side. The procedure used in experimenting upon this, as upon the other kinds of transfer to be discussed presently, is of the following form, when L represents one or more practice trials:

Control group: L_1	No further training	L_1 (continued)
Experimental group: L_1	Training on L_2	L_1 (continued)

The two groups are equated as nearly as possible at the beginning. The brief introductory practice L_1 gives a measure of performance prior to training, and a comparison of the L_1 (continued) scores of the two groups yields a measure of the transfer from the training on L_2. Bilateral transfer has been found in a large number of acts such as mirror drawing, speed of tapping, ball tossing and maze running.

Transfer from one class of material to another. The next question is whether practice upon one or more samples of one class will facilitate the learning of samples of a different class. Will practice at learning nonsense syllables transfer to learning poetry, or will practice at maze learning facilitate mirror drawing? This question refers to a third form of transfer of training.

The experimental answer to this problem is that any one of three possibilities may take place. (1) Practice upon one class of material may assist the learning of a different class, *i.e.,* there may be positive transfer. (2) Practice upon one class may have no measurable effect upon the learning of a second class. (3) Practice upon one may hinder the learning of a second, *i.e.,* there may be negative transfer. The specific conditions of the *situation* determine which one shall occur. The large amount of experimental work upon the problem permits of no doubt that each occurs, although under most laboratory conditions positive transfer is the most common. It can safely be said that *every instance of learning is influenced by the already existent associative organization of the learner.* The experimental problem is to discover the conditions which determine which aspects of the organization will exert this influence and in what amounts.

Determining conditions of transfer. We are still seeking for information so precise that we can understand why a given case of transfer has occurred and can predict the occurrence of other cases. The two chief possibilities in respect of which research is planned are the following.

(1) Transfer may take place in terms of specific elements

which are identical in (common to) the activities involved. The identical element may belong to the material or to the response. Practice at addition transfers to multiplication because the latter involves addition. The identity may, on the other hand, be one of method. Practice at one maze may facilitate the learning of a second maze because the method of attack is the same in both. One difficulty with this as a general theory of transfer is that there is seldom a complete identity, but rather a relatively high degree of similarity. Transfer in terms of this similarity sometimes approaches a generality which fits better the second view, to which we may now turn.

(2) Transfer may take place by means of the generalization of principles, methods or motives. Thus, the theory of evolution, the scientific method or an ideal of neatness, learned in one context, may operate to aid learning in a very different context. Transfer of this kind certainly occurs, but it is not automatic or inevitable. The generalization must be acquired as one which may be applicable to more than the context in which it is acquired or it is unlikely to transfer. Witness the scientist who is a consistent adherent to the scientific method in his own field, but who forsakes it completely when he deals with philosophical or theological problems.

It may be suggested that these two theories of transfer are not necessarily opposed. They may be viewed as representing different parts of a continuum reaching from a high degree of specificity at one end to a high degree of generality at the other.

Associative inhibition. One of the conditions of negative transfer is stated in a generalization known as the *Muller-Schumann law* which reads thus: *When any two items, as A and B, have been associated, it is more difficult to form an association between either one and a third item, K.* The inhibition of *A-K* by the prior learning of *A-B* is more pronounced when *A-B* has been learned to only a moderate degree. An example of the phenomenon is afforded by the greater difficulty of learning that Jones's telephone number is 4916 when one has learned only fairly well his last year's number, 8921.

The gross fact is illustrated, also, by many phenomena of *habit interference*, such as are found in changing to a car with a different gear shift or in learning a maze in which the correct turns are the opposite of those in a maze previously learned.

DESCRIPTIVE THEORIES OF LEARNING

The problem. We have been dealing with some of the conditions which determine rate of learning. We have set out, in describing them, from the fundamental fact that phenomena called learning occur. These phenomena and the conditions which we have described are facts and remain so, in spite of any theoretical arguments about their interpretation. It is important, however, for the construction of an organized body of knowledge, to arrive at a more general understanding of these phenomena. To this end we shall review three of the major views concerning the nature of learning, that is to say, concerning what the increments of performance (which we call learning) involve and how they may most usefully be described. This section is introductory to the next one, which deals with the more general conditions of learning.

Learning as a conditioned response. Learning is often described as the formation of conditioned responses. This view had its origin in a famous experiment performed by Pavlov upon dogs. Having first made sure that the visual perception of food would elicit a flow of saliva and that the sound of a bell would not, he presented bell and food with some temporal overlap or in immediate succession for a number of times and found that presentation of the bell alone would then cause salivation. Everyone whose mouth has 'watered' at the odor or verbal description of food has exhibited a similar phenomenon. An analogous technique has since been used with human subjects, and it has been found that a large number of responses can be conditioned in the same general way.

The stimulus which was already connected with the desired response at the beginning of the experiment is said to be the

primary stimulus, the one to be substituted for it is called the *conditioned* or *secondary stimulus* and the response to the latter is called a *conditioned response*. It should be emphasized that the conditioned response is not *identical* with the primary or unconditioned response. The former is at most but one aspect of the latter and may differ from it in kind to a considerable degree. The salivary response of the dog is but one part of the original pattern of response elicited by the presentation of food, *e.g.*, saliva flows and the dog leans forward. The remainder of this pattern may and often does fail to appear when the secondary stimulus is given. Instead a different pattern may be present, *i.e.*, saliva flows and the dog stands still. The two patterns are different, but salivation is common to both. The secondary stimulus does not, therefore, act as a substitute for the primary in calling forth the *same* response. Rather, the secondary stimulus, as a result of its paired presentation with the primary, has acquired the property of eliciting at least one aspect of the response made to the primary or a response in some important way related to the primary response. (For further description of the conditioned response see pp. 432, 434-436.)

It is the belief of some psychologists that the conditioned response is the prototype of all learning. It seems, however, that although new experimental methods have been derived from the experiments upon conditioned responses and although many new facts have been gathered, no fundamentally new principle has been introduced. The conditioned response is in principle the same as association, which we discuss in the next paragraph.

Association theories. A second view of learning describes it as an *association* between two terms. When one learns that *Gedächtnis* means *memory*, one is said to associate the two terms so that the presence of one as stimulus or instigator will elicit the other as response or instigated item. The learning of serial performances or materials is then described as the formation of associations between the constituent parts, whether

they are correctly presented to the learner at once or discovered by trial-and-error, so that the parts become organized into a single pattern. We have already seen that, during practice, associations are formed between spatially remote items in both the forward and backward directions as well as between adjacent items. This produces a complex weaving together of the parts into a more compactly organized whole than the connection only of adjacent items with each other could give.

Gestalt theories. The objection has been urged, however, that association implies a large number of chaotic and disorganized items or events which, under proper conditions, become associated, and that mental life does not get organized in this way. Rather is it true that mentality is organized at the outset. What we call learning is a differentiation and complication of already existent organization. It implies development of *insight* or of grasp of the relationships involved, as when one suddenly 'sees' how to solve a puzzle. Association is merely an outcome of this change in organization, not the means of its attainment. The Gestalt psychologists, who hold this view, insist that only on such an assumption can organization be understood. The association view that organization is effected by an association of originally unconnected parts or atoms is, they say, impossible because there is no known principle by which such atoms could become connected.

A discussion of association and Gestalt theories. It will be noted that the definition of association given above asserts only that association is a term descriptive of a fact. Association does nothing. Whether it occurs or not is a function of the conditions obtaining at the time. In the Gestalt view, likewise, organization and insight are descriptive terms applicable to facts. The issue between the two views hinges on the question as to which is the better description. This is not the place for a detailed examination of these theories. It should be said, however, that at the present time few, if any, adherents of associationism conceive of the associated terms as isolated

atoms. Even the most meaningless nonsense syllable has, for a subject of any but the earliest chronological and mental age, a context of meaning. When two such syllables are associated, the context of each acts as a condition of the association and to some extent enters into connection as well as the syllable itself. Two wholes, to use a Gestalt term, are, thus, connected. That which is described as insight in Gestalt theory, may be thought of as transfer or associative spread.

(1) The associationist sees learning as the connection between two different wholes, whereas the Gestalt psychologist regards it as the formation of a single new whole. (2) For the one, association is the primary descriptive concept; for the other, it is a derived and secondary concept. Which view one adopts will determine one's interpretations of the facts of learning, but it cannot alter the facts themselves. Since most of the experimentation has been done from the associative point of view we have employed its terminology.

GENERAL CONDITIONS OF LEARNING

The problem. The existence of an organism which practices and of a material or performance which is practiced are necessary if learning is to occur at all. We have dealt with some of the major characteristics of the learner and of the thing learned in relation to rate of learning. There remain, however, certain general conditions which must be satisfied if any individual is to learn any material or activity, and in dealing with these conditions we shall use the concept of association with the meaning given to it in the preceding section.

The temporal conditions of learned events. A certain degree of temporal proximity is necessary for the formation of an association between two items or events. Neither simultaneity nor immediate succession is, however, necessary. Associations may be formed between terms which are separated by an interval at least as long as several seconds and sometimes by a much longer time. The fact that a certain degree

of temporal proximity is a necessary condition of association is known as the *law of contiguity.*

Associative spread. For learning to occur at all, the subject must be able to perceive and respond to the material to be learned. He is able to do this because of the fact that his already existent organization carries over to or spreads to the new material. The retained effects of prior practice are to this extent general. As a consequence the material to be learned is at the beginning of practice partially fitted into the old, and during practice there is a mutual interaction of old organization and practiced new material, until finally the new has been assimilated. If everything learned were entirely specific and never related to new material, learning as we know it would be impossible.

The phenomena of transfer, which we discussed earlier, are special cases of associative spread. The fact of spread is illustrated whenever one perceives similarities or makes associations by similarity or, for that matter, whenever one perceives or learns anything. By means of it the new is assimilated and the complex pyramiding of new-on-old, which learning represents, is made possible.

Motive and effect. Practice assumes motivating conditions in the subject; otherwise he would not practice. These motivating conditions are complex and may include many conditions other than those controlled by the experimenter. Whether or not learning occurs is a function of the *effect,* as it is called, of the successive practice acts. Effect refers to the knowledge of the results of the act and to whatever consequences follow it, and is always relative to the motive or motives operating at the time. In the absence of such consequences learning occurs very slowly, probably not at all. Other things being equal, those acts which lead to consequences which satisfy a motivating condition are selected and fixated, and those which lead to consequences which do not satisfy a motivating condition are eliminated or inhibited. This is the *law of effect.* The influence of effect is basic in accounting for

the selection of some acts and the elimination of others in the learning of skills and the solution of rational problems.

The complexity of motives, and especially of those which the experimenter is unable to control, often makes it difficult to say whether a given instance of learning has satisfied a motivating condition. *Incidental learning,* the learning of casually observed items or of casually performed acts, seems to be unmotivated. Motives and effect exhibit, however, a characteristic analogous to associative spread, that is to say, they apply to at least part of the context of the specific act. It is reasonable that apparently incidental learning is explicable in terms of motivation uncontrolled by the experimenter and of the spread of association and effect.

The objection has been raised that the operation of effect implies the reversibility of time in that consequences of an act work backward to select and fixate an act already performed. The implication is not necessary. Effect works forward and may operate to influence the next performance of the act, to remove inhibitions of that performance or to do both.

Frequency. The things which human beings learn, if of any degree of complexity, require at least several repetitions; everything learned requires at least one. It has often been assumed that frequency of repetition, in and of itself, selects and fixates. This conception deserves critical examination.

1. This concept has usually implied that successive repetitions of any act or material are repetitions of the same thing. Actually, however, successive repetitions are far from being identical. In learning a maze each run from start to goal may have different spatial and temporal characteristics, and, even after the errors have been eliminated and the true path selected and fixated, successive correct runs are not identical. The subject may go down one side of the path on one run, down the other side at the next, and may cross from one side to the other in the middle of a section on the third. Variability of performance throughout learning is the rule. The pattern to be learned is approached by a series of successive approxima-

tions. This being the case, what is repeated? The attempt to get from start to goal, or to carry through whatever the conditions demand, is the only thing repeated. The specific performances whereby this attempt is carried out may vary widely. There can be, therefore, no stamping-in of specific neural paths by repetition, as the older frequency theories assumed.

2. Repetition does not necessarily select and fixate. It may also inhibit and eliminate. The making of an error in maze learning or in solving a rational problem aids elimination of the error if elimination satisfies the prevailing motivation. It is thus that errors in typing, nail-biting and other acts are inhibited by repetition when repetition is used intentionally. This finding seems to be a special case of the law of effect.

3. In a long series of experiments, Thorndike has shown that frequency of a connection or act, even when successive repetitions are as alike as they could be made, yields little if any learning. When he added to frequency a factor which he called *belongingness*, by which he meant that the frequently connected terms seemed naturally or rightly to go together, somewhat more evidence of learning appeared, but still not much. Belongingness involves transfer or spread from prior learning and perhaps from other motives than those supposedly aroused by the experimental situation and thus goes beyond mere frequency. It is very doubtful that frequency by itself is a determiner of learning.

4. The empirical correlation between practice and the increments of performance which are learning cannot, therefore, be referred to the repetition itself. The correlation results from the *effect* which repetition leads to and from the associative *spread* which it permits to operate. Frequency acts as a carrier of *effect* and *spread.*

RETENTION AS A FUNCTION OF TIME

Measurements of retention. We have been dealing thus far with the course of the learning process and with the conditions which affect it. This learning, as we noted at the beginning of the chapter, necessarily involves a progres-

sive retention throughout its course. Experimentally, however, we speak of measuring retention only when we measure learned performance at some time after a given criterial level has been reached. We then inquire how much of the criterial performance will appear under measurement after the interval. Theoretically, we infer the retention of some neural correlates of the learning and we also infer that there may be more of such retention than appears under any given measurement; but experimentally we define retention as performance under certain conditions when compared with performance prior to the elapsed interval.

The numerous ways in which retention may be measured can be grouped under four headings: (1) *reproduction* (or recall or reinstatement), in which the subject is placed again in the original stimulating situation, confronted with the same problem, or given the same cue, and is asked to perform as nearly as possible as he did at the criterial level; (2) *saving,* in which we require that the subject reach the criterial level obtained during learning, after which we compute the difference between the learning and the relearning scores and express it as a percentage saved; (3) *recognition,* in which the subject is confronted with the material learned, mixed at random with similar items, and is asked to indicate those which had been in the original material; and (4) *reconstruction,* in which the material learned is presented in a new arrangement and the subject is to rearrange it in the order in which it was learned.

The form of retention curves. In general we wish to know how much of any given performance is retained, not only after a single interval, but also after as many different intervals as possible. By taking the proper measurements of performance after a number of intervals, we obtain the data for a curve of retention. This measurement was made first in 1885 by Ebbinghaus, whose work is one of the classics in the field of learning. He employed the method of saving with nonsense syllables learned to a criterion of two perfect repeti-

tions and obtained the curve shown in Fig. 111. This curve has a negatively accelerated form, falling rapidly at first and then at a steadily decreasing rate. Most curves which represent measurements taken within relatively short intervals after the end of the learning period have this general form. The specific amounts retained vary with a number of conditions presently to be discussed, but under most conditions retention decreases rapidly during the time immediately after learning and thereafter at a decreasing rate.

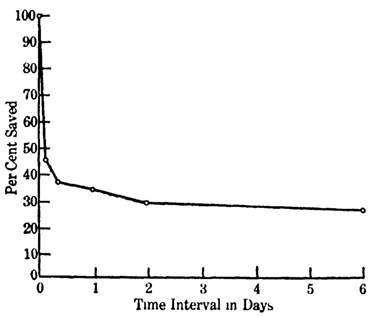


FIG. 111. EBBINGHAUS' CURVE OF RETENTION AS MEASURED BY THE SAVING METHOD

From H. Ebbinghaus, *Memory* (tr. Ruger and Bussenius), New York: Teachers College, 1913, p. 76. The curve is plotted from the data given there.

Reminiscence. A few striking exceptions to the negatively accelerated retention curve have appeared under conditions which make the exceptions seem valid. When the method of reproduction is employed, particularly with meaningful material, and only with material incompletely learned before the interval, a large percentage of the persons in an experimental group may reproduce somewhat more items after an interval of from one to two days than immediately after learning. The curve for these subjects, and sometimes that representing the average of the whole group, may then rise over the first day or two, falling thereafter at a negatively accelerated rate. This phenomenon of better reproduction at some time after practice than immediately after is termed *reminiscence*. Its conditions are very imperfectly known, and we note here the fact of its occurrence without speculating upon what determines it.

CONDITIONS OF RETENTION

Method of measurement. The relative amounts of retention are a function of the method by which the measure-

ment is made. In general the amount of indicated retention is greatest by the recognition method, less by the reconstruction method and least by the reproduction method. The saving method commonly yields relatively smaller amounts of retention than do at least some of the others after short intervals and relatively larger amounts after longer intervals. These facts are represented in the curves of Fig. 112. Each of these methods measures what we call retention, but each one provides a different stimulating and instructional context with

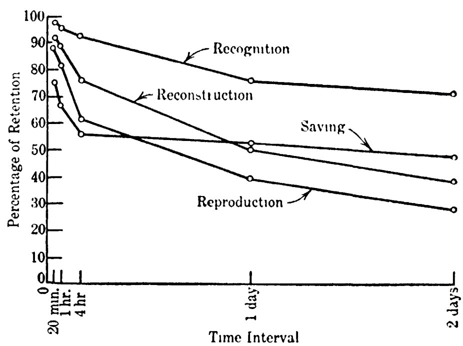


FIG. 112. RETENTION CURVES YIELDED BY DIFFERENT METHODS OF MEASUREMENT. From C. W. Luh, *Psychol. Monog.*, 1922, 31, No. 142, p. 22

consequent differences in results. Generalizations about retention are, of course, specific to the operations by which the performance has been measured.

The characteristics and behavior of the learner. Since retention so completely pervades learning, we should, other things being equal, expect conditions to affect retention after an interval in essentially the same way as they affect it during learning, and to affect relearning as they do learning. To a large extent this expectation is justified. (1) Measurements of retention yield an age-curve which resembles in form that for age and learning. (2) Absolute amounts retained are positively correlated with intelligence. Whether relative amounts are correlated in the same

way is not known. (3) Learning with intent to remember indefinitely has a favorable influence upon retention. (4) Recitation during learning leads to an absolutely higher degree of retention. (5) The range of individual differences in retention is of the same order as the range during learning.

The characteristics of the material and mode of presentation. (1) Meaningful material is much better retained than non-meaningful. A comparison of Figs. 111 and 113 will illustrate this fact. (2) Retention increases within fairly wide limits with the amount of the material learned. (3) Distribution of practice during learning is superior to massed practice for retention as well as for learning, and the duration of the practice periods has been shown with some materials to be a more important variable for retention than is length of rest interval. (4) Although measurements of degree of learning and retention are necessarily functions of specific methods of measurement, in general we may say that retention increases as does degree of learning until a considerable degree of overlearning is reached, after which additional repetitions of material already mastered may have an inhibitory effect.

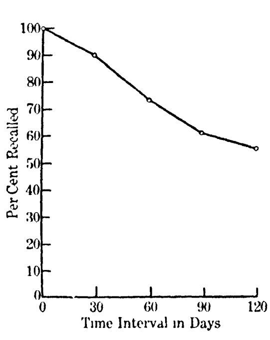


FIG. 113 A CURVE OF RETENTION FOR MATERIAL (OBJECTS) BRIEFLY OBSERVED

From J. A. McGeoch and P. L. Whitely, *J. Educ. Psychol.*, 1926, 17, p. 422.

More general conditions of retention. We have been discussing the relation of the characteristics of learner and of thing learned to the retention of the latter after an interval of no practice. In one sense these conditions are general in that they provide the necessary framework or setting for both learning and retention. They do not, however, state the conditions which operate to decrease performance below its level at the close of the learning period, and which must operate if

such a decrement is to occur. We shall now turn to the experimental work on these more general conditions.

It is necessary, first of all, to consider the frequently mentioned view that forgetting is to be accounted for by increased resistance (or other changes) in the nerve pathways. It is the organism which learns and retains, but we do not at present know what specific changes in the organism are correlated with these aspects of performance. We do have a large amount of factual information about performance and its conditions. It is more fruitful to attempt to account for our phenomena in terms of conditions which we are able to control experimentally than in terms of speculations about what may be going on in the organism. We shall, therefore, continue this discussion at the phenomenal level.

The influence of interpolated activity. It was demonstrated early in the history of research on learning that learning another sample of the same material during the interval between the end of practice and the measurement of retention was followed by a decrement in retention. The phenomenon has since been found under a large number of conditions. The name *retroactive inhibition* is used to designate it and to refer to any decrement in retention produced by activity interpolated between the end of practice and the measurement of retention.

The experimental procedure employed in studying retroactive inhibition is of the following general form, when the numbers stand for learning materials of whatever kind:

Rest condition .	Learn (1)	Rest	Measure retention (1)
Work condition .	Learn (1)	Learn (2)	Measure retention (1)

The difference between the performance at the time of measuring retention under the two conditions is the gross amount of inhibition, and this difference divided by the measure under the rest condition is the relative amount. The two conditions are arranged so that they differ importantly only in respect of the interpolated period (rest, or learning) between the original learning and the measurement of its retention.

The experimentally obtained decrements from interpolated learning vary in amount from near zero to near 100 per cent. The amounts which appear are a function of a number of conditions. Meaningful material is less susceptible to retroactive inhibition than are lists of disconnected items. Over a considerable range of similarity, degree of inhibition varies directly with the degree of the similarity. It also varies inversely as the amount of practice given the original material. The longer the material and the more difficult it is to learn, the less susceptible it is to inhibition. Inhibition may also be decreased by instructions to the subject to try to avoid interference from the interpolated learning.

The experimental data upon the inhibitory effects of interpolated learning show that a large decrement in retention can be produced by such interpolation. They imply, also, that such decrements are a matter, not of passive decay or disintegration with time, but of active blocking of one performance by another. The facts thus far cited do not, however, demonstrate that blocking of this kind will account for all the forgetting which occurs in the course of everyday life. There are, however, other experiments which have this implication. Two of these researches, in particular, are of the first importance for this problem. These experiments have studied retention after equal intervals of sleep and waking and have found a relatively large amount of superiority for the sleep condition after intervals longer than an hour. The difference between the results under the two conditions may most reasonably be referred to the events which filled the intervals of waking. These events correspond to the interpolated activity of the experiments first described, while the intervals of sleeping correspond to the rest condition except for the fact that sleep is much less filled with activity than waking 'rest.'

The available facts strongly support the interpretation that the decrements in retention which we commonly find correlated with the passage of time are a function of the events which fill the time, that is to say, retroactive inhibition is one of the primary conditions of forgetting. In the course of daily

life each new experience or response is an interpolated activity with reference to what has gone before. Short of a psychological vacuum, opportunities for such inhibition cannot be escaped. A large amount of decrement can be avoided only by a favorable operation of the factors which determine degree of inhibition.

Altered stimulating conditions. Everything which the human being learns has, during practice, a complex context which is a part of the total stimulus situation. We may call it the stimulational context, and it may be subdivided into three groups. One kind consists of the stimulation from the external environment, such as the furniture of the room, the experimenter and the apparatus. A second kind of context is the stimuli from the interoceptors, and a third is the ideational context present. These factors may be, and many of them are, connected with the material learned. These contexts alter with time, and the alteration creates a situation in which, although the intent to recall is present, the necessary eliciting stimuli are not. An example of this situation is seen in the failure to recall the name of a person seen in a new environment, although the name is readily reproduced in the environment customarily associated with the person.

Some stimulus is always necessary to elicit a learned performance. It is probable that the failure of a previously learned performance to appear may often be referred to the fact that the stimulational context necessary to elicit it is not present. This condition, together with the inhibitory effect of interpolated activities, offers our most reasonable explanation of the phenomena of forgetting.

PERVASIVENESS OF LEARNING

The fact of learning is one of the most indubitable and universal of the phenomena of psychology. It can be treated as a phenomenon separate from other aspects of mental life only by arbitrarily isolating it from them. It is necessary, for the purposes of experimentation, to abstract learning and reten-

tion from the rest of mental activity and to treat them as relatively independent phenomena. Actually, however, they occur in a context of other events, each of which must likewise be abstracted if it is to be studied. Learning and retention intimately involve sensory stimulation and perception, movement and action, thought and affectivity. The learning of an act of skill, such as a maze habit or swimming, is a matter of organizing the perceptions of spatial and temporal relations into a new pattern. However, the perceptions and movements thus organized were for the most part learned at an earlier time. In the same manner, the learning of a verbal series, like a poem, involves perception of rhythm and meaning, an hedonic tone, the use of representative functions, the organization of articulatory and related movements into a new sequence. Each of these factors has, in turn, been brought by practice to the status it occupied at the beginning of the learning of the poem.

Learning is present in and pervades all but the most elementary phenomena of mental activity, and they, conversely, are present in learning. The conditions which determine learning, however, are the conditions under which the organization within and among the other phenomena occurs. These conditions are, thus, the conditions of mental organization. It is no wonder that the capacity to learn has been proposed as the criterion for the existence of mind.

REFERENCES

1. CARR, H. A. *Psychology: a study of mental activity.* Chicago: Longmans, Green and Co., 1925. Chaps. 5 and 9 to 11.
2. CARR, H. A. Teaching and learning. *J. Genet. Psychol.*, 1930, 37, 189-219.
3. CARR, H. A. The laws of association. *Psychol. Rev.*, 1931, 38, 212-228.
4. HUNTER, W. S. Experimental studies of learning. In *A handbook of general experimental psychology.* Worcester, Mass.: Clark University Press, 1934. Chap. 11.
5. MCGEOCH, J. A. Experimental studies of memory. In *Readings in general psychology* (edited by E. S. Robinson and F. Richardson-Robinson). Chicago: University of Chicago Press, 1929. Pp. 369-412.
6. MCGEOCH, J. A. The psychology of human learning: a bibliography. *Psychol. Bull.*, 1933, 30, 1-62.

7. PETERSON, J. Learning in children. In *A handbook of child psychology.* Worcester, Mass.: Clark University Press, 1933. Chap. 10.
8. ROBINSON, E. S. *Association theory to-day.* New York: D. Appleton-Century Co., 1932.
9. THORNDIKE, E. L. *Human learning.* New York: D. Appleton-Century Co., 1931.
10. WHEELER, R. H., AND PERKINS, F. T. *Principles of mental development* New York T. Y. Crowell Co, 1932. Chaps. 5 to 7 and 13 to 22.

CHAPTER 14

IMAGERY

In the preceding chapter the effects of past experience on memory and on skill in manual habits have been discussed. After repeated experience in learning a list of words or running through a maze, the behavior of a person changes; he can repeat the list or run through the maze more accurately. This, however, is not the only effect of past experience. After a stimulus has ceased to act and the perception of it has died away, something similar to the perception can be recalled. One experiences the perception, or part of it, again, despite the fact that the stimulus is no longer present. Thus, if one thinks of an absent friend, there often occurs a memory of the appearance of his face or the sound of his voice. If one recalls a recent automobile ride, it is possible to see in the 'mind's eye' the appearance of the front of the car, to hear again the roar of the motor or to feel the cold draft of air from the window. These experiences are partial reproductions of the perceptions of the past. Conscious memories of this sort, which reproduce a previous perception, in whole or in part, in the absence of the original stimulus to the perception, are known as *images*.

NATURE OF IMAGES

Images make up a large part of our mental life. Hence one finds references to them throughout popular and scientific literature. The poet Goethe reported that when he closed his eyes and thought of a rose he saw a sort of rosette, of constantly changing appearance, with green or red petals un-

This chapter was written by C. W. Bray of Princeton University.

folding from within. The psychologist, William James, stated that a student of his, who often learned poetry by heart, could, after memorizing a page, see a visual image of the first words of each line. As an example he gave the following from a poem by La Fontaine:

> Etant fait . . .
> Tous . . .
> A des . . .
> Que fit . . .
> Cérés . . .
>   Avec . . .
> Une fleur . . .
>   Comme . . .

When he thought of the poem, these words were seen in visual imagery. The words at the beginning of the indented lines were less clear than the others in the image.

It is frequently thought that the musician Beethoven must have had very strong and vivid auditory imagery, for he composed several long symphonies after he became completely deaf. By using imagery he could play over the score of his composition and thus get the sound of the music even though he could not hear. Although this probably was the method he used, it is seldom safe to draw conclusions about imagery from casual accounts of the achievements of a great man.

Only careful introspection can distinguish between the various forms in which memory occurs. A clear and accurate description of past events can be made without any imagery or mental reproduction of the events at all. Thus, in describing the appearance of one's home from memory, one can name the style of architecture, outline the yard around it and state the size of the garage without actually having an image of the home. In this case verbal habit takes the place of imagery. One does not 'see' the home as it has actually been seen before.

Therefore, in dealing with the subject of imagery, one must be careful to distinguish between knowledge about an object, which is generally carried in verbal terms, and ability to re-

produce the object in imagery. The two are not always the same, nor are they necessarily even related to each other.

The term image suggests that reproductions of perception are very complete, as in photographic pictures of scenes, and

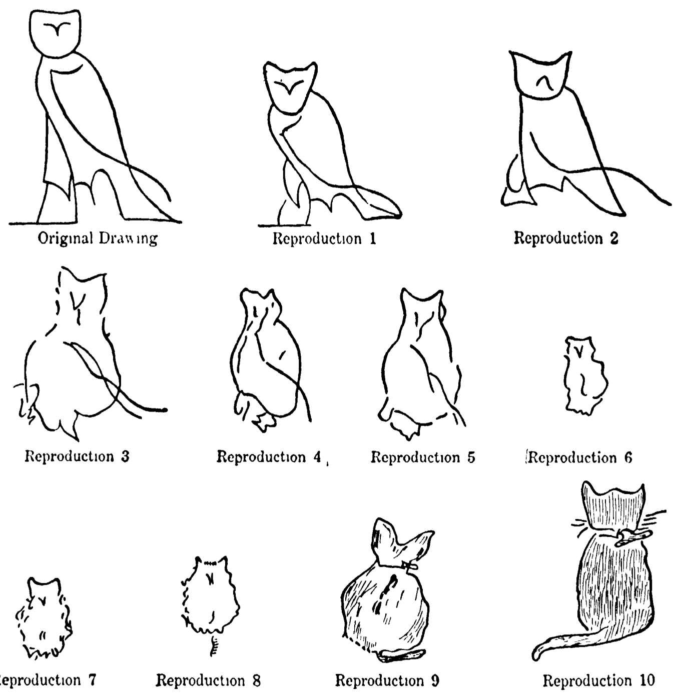


FIG. 114.

The owl-shaped figure of the Egyptian letter 'mu' in the upper left-hand corner was shown to a person. His reproductions of the figure, made at successive later periods, are given in the following diagrams The reproductions show a progressive modification of the memory from period to period. From Bartlett (2).

that images are solely visual. Neither supposition is true. Images are seldom exact copies of past experience. Even in relatively complete images some details are almost certain to be omitted or false details added. In a way, imagery (and all memory as well) is *creative*, rather than imitative, for the process of imagination involves an active reaction and is not

merely a passive recall of past events. The image is influenced by the person's interests and attitudes and by past experiences (2). In this, imagination is like perception, which, as pointed out in an earlier chapter, is not a process of passive reception of stimuli but a process involving an active integrating reaction to stimuli. Just as a perception is never a mere copy of a stimulus, so the image is never an exact copy of a perception. An example of change in an image is shown in Fig. 114. The drawing at the top left was shown to a person; the subsequent drawings are his reproductions made from memory at successive later intervals. The original owl-shaped figure was changed from drawing to drawing until it resembled a cat. In other words, the subject remembered the original figure as a cat and not as an owl.

It is also a mistake to consider that imagery is always visual in character. Visual imagery is prominent in mental life, but auditory, tactual, organic and other images are experienced as well. It is safe to say that all the senses, with one possible exception, to be discussed below, contribute to imagery. In fact most images are *composite*, involving components from many senses at one time. In this characteristic, also, images are like perceptions, which are formed of data from many senses simultaneously. The dominant or clearest component serves to classify an image as of visual, auditory or other type.

The possible exception to the statement that all the senses furnish images is kinesthesis. It is doubtful whether true images of our movements ever occur, for it is thought that when we remember a previous movement there is a tendency to move the muscles slightly. Consequently, kinesthetic memories are believed to be true perceptions of actual movement rather than imaginal reproductions of previous perceptions. Evidence bearing on this belief is derived from both introspective and objective studies of thought and action. For example, when one recalls the movements of running and dodging away from another person or the movements of a golf or tennis stroke, it is possible to observe, introspectively, strains and other sensations arising from the muscles of the body. Sometimes the

movements evoking these sensations are so gross that the body can be seen to sway and move. Even when the movements are so fine that they cannot be observed with the eye, electrical methods of observation confirm the introspective evidence of actual movement (see pp. 481f.).

Strictly speaking, then, memories of movement are not images because these memories occur in the presence, rather than in the absence, of the stimuli to a perception of movement of the body. Nevertheless memories of movement have usually been called *motor* or *kinesthetic images.* This terminology is not exact, but it may be followed since motor memories are similar in function to true images and are evoked by the same conditions.

Introspection reveals that memories involving language are frequently expressed in movement of the vocal muscles in the throat and mouth. The movements are very weak and usually no sound is made so that they pass unobserved unless attention is directed to them. These movements are known as *vocal-motor images.*

One remembers not only the words spoken by oneself, but also the words of others, in terms of vocal-motor imagery. Music and noises, likewise, may be recalled in this manner. There is a very close connection between speech and hearing, due to the fact that, when one speaks or makes any other sound, one almost always hears this sound. Auditory memories, then, may involve vocal-motor imagery as well as simple auditory imagery.

We have seen that memory can be expressed in true imagery and in motor imagery. It can also be expressed in overt or strong movement. This difference in the ways of expression is excellently illustrated in an introspective account by Titchener (18) of his methods of reading to himself and lecturing to a class. In reading he found that the printed words were spoken subvocally, *i.e.,* the sight of the words gave rise to vocal-motor imagery. The movements of articulation were brief and reduced to a minimum, but on occasion, with prolonged reading, his throat would begin to feel stiff as with too much talking. In

lecturing, Titchener observed that, when he wished to follow a set procedure or definite organization of the material to be given, he would refer to a visual image of the manuscript or outline of the lecture. The image helped to orient him and to keep the material in its proper place in the lecture. At other times, when the subject-matter was difficult, he tended to think ahead to the coming words. These he experienced in auditory imagery. While saying certain words aloud, he was at the same time imagining others. This procedure, as he pointed out, sounds confusing, and yet is not actually so. When the material was easy and familiar to him, he let his voice take care of itself and talked from habit, paying attention to blackboard illustrations and the like.

THE ROLE OF IMAGERY IN MENTAL LIFE

Imagery and problem solving. The various kinds of images occur in connection with many activities of the individual. In perception, memory, emotion and nearly all mental activity, images arise as a part of one's reaction to stimulation. As a general rule it can be said that images occur in greatest number when the *situation does not call for action, or when movement is difficult*, and that few images are experienced when a stimulus calls out an unimpeded response (8). A flood of imagery arises when an individual is faced with a problem, that is to say, with a relatively new situation for which he has no habitual response. Also when stimuli call out conflicting reactions, when doubt arises as to a course of action, then many images are experienced. These images represent possible ways of solving the problem. They symbolize modes of action or facts previously learned, which might be helpful.

It is partly by means of images that past experience is brought to bear on the situation. Thus in walking home over a very familiar road no imagery of the route occurs. But in walking over a somewhat unfamiliar road, images of landmarks, of familiar objects seen a moment before and of one's own previous movements furnish clues for orientation. One thinks

of going toward a building, of turning right or left, of retracing one's footsteps. These are tentative, mental solutions of the problem of finding one's way. Mental trial-and-error precedes overt trial-and-error, and the probable success or failure of any trial is weighed consciously before it is undertaken (see pp. 475f.). In this process each trial and its probable consequences, so far as these can be determined from past experience, may be represented in imagery.

The role of imagery in the solution of problems is often seen even in idle fancies, imaginative play, daydreams and the dreams of sleep where no problem is apparent to the person experiencing the imagery. Children who are left too much alone invent an imaginary playmate to help fill their solitary hours, even though they do not realize their need for companionship. When they are mistreated by a servant or teacher, the images of their dreams and daydreams represent ways of getting rid of the offending person. Often the only way to obtain insight into the needs of children is to analyze carefully the meaning of the imagination images in which they indulge.

In the daydreams and fancies of the adult the same fact is sometimes seen. Castles in Spain, dreams of riches or of superlative skill in work or play, are imaginary solutions of the problems of everyday life. Even the bizarre images of dreams, as Freud has shown, can sometimes be related to the unfulfilled desires of the dreamer (9). An example is that of a university teacher, who, feeling vaguely that his work was misdirected and meaningless, dreamed of a squirrel running around and around a room, fruitlessly jumping here and there in an effort to escape. At last the squirrel disappeared in a mysterious manner and was felt to have escaped. When questioned, the teacher stated that the situation of the squirrel reminded him of his own dilemma from which he could escape only by a miracle.

Although images are prominent in the mental life of most individuals when problems and conflicts occur, it does not

follow that they are necessary to the solution of problems in all individuals. Some individuals state that they have never experienced imagery of any sort, except perhaps the motor imagery of subvocal speech. These individuals may attain considerable success in fields in which it would seem that the possession of imagery would almost be a necessity. Thus, there are the cases of a painter and of a geometrician, each said to be quite successful in his field, who never experienced visual imagery. And there have been experienced musicians who have had little or no auditory imagery and who have had to compose by trying out successions of notes on an instrument.

The experimental investigations of the role of imagery in the solution of problems have been carried out by giving individuals problems to be solved mentally. After completing the task, the subjects describe the nature of the imagery used. With some problems it is apparent that clear imagery, if suited to the task, aids considerably in the solution (6). Thus the following problem is most readily solved by those having a clear visual image.

Imagine a 3-inch cube painted on all sides. If this cube were divided into smaller cubes of 1 inch each, how many would have paint on three sides; how many on two sides; how many on one; and how many on none of the sides?

In this problem a clear visual image of the large cube and its subdivisions is of great value. Motor imagery also helps some subjects, who 'pick up' the smaller cubes and turn them over, looking at each side. Subjects with scanty visual imagery are slower in solving the problem. On the other hand the problem can be solved entirely by mathematical means, that is by vocal-motor imagery alone. And, further, individuals find at times that a clear visual image unsuited to the problem may arise and block its solution. The face of one's beloved may come between one and one's work!

Experience is an important factor in this type of mental work. This fact is implied in the statement that the image must be suited to the task if it is to aid in the solution. And

the effect of practice is shown if, after solving the 3-inch-cube problem mentioned above, one tries to solve the related, but more difficult, problem of dividing a 4-inch cube into 1-inch cubes and calculating the number of painted sides on each.

Images and associated ideas. Not all images, by any means, are reactions to problems. Many arise as simple *associations.* In reading, the words of the printed page stimulate associated imagery so that, for example, when one reads a literary description of a storm, fragments of one's past experiences with storms may be recalled. Memories of clouds and lightning, thunder, rain and of running to shelter are associated with the verbal stimuli describing these events. In revery, a stimulus sets off a whole train of associated ideas, each image in the train leading to the next. These images are not necessarily due to any problem or unfulfilled wish of the individual. Their character is largely determined by past experience, that is to say, by their association with stimuli of the type evoking them.

Images and meaning. As we observed earlier in the chapter, images are not photographic. Usually only a small part, sometimes only a minor detail, of a previous experience recurs in the image. The response, which solves the problem of finding one's way home over a somewhat unfamiliar road, may be represented in consciousness by an apparently irrelevant detail of a past experience, such as the color of a building which one must approach. Yet this absurd image stands for an action or object which is closely related to the problem. The image has a *meaning* far more important than its form indicates. In this it is like a word, which, objectively, is only a series of black marks on a page, but which, subjectively, may have a very important meaning.

Images carry meaning by virtue of their association with previous events and actions. An infant soon associates the sights and sounds made by his mother with the movements he makes in eating. He responds to the sight of the mother with sucking or eating movements whether or not she feeds him

at the time. Apparently she means food to him. As he grows older the overt sucking and eating movements are inhibited and suppressed, unless food is actually present. Stimuli associated with food now call out motor images of eating and memory images of previous situations in which he has been fed. These images may be fragmentary or partial reproductions of the previous experiences, as for instance, a visual image of a white dress of the mother or an auditory image of certain words spoken by her. The images stand for and mean the reaction of eating food. They take the place of the overt movements of eating when the child thinks about food. When he is older and hears a dinner bell, the apparently unrelated image of the white color of the dress is experienced and adds meaning to the sound of the bell.

The meaning, not only of actions, but also of words and other stimuli, is often given by imagery. The images which are aroused by reading a description of a storm give meaning to the words read. A symphony concert stimulates images associated in emotional mood with the music, and these images give meaning to the sounds made by the orchestra. The fact that the imagery is often fragmentary or even absurd does not detract from the value of the experience, for, by association, the image stands for many previous reactions and is felt to take the place of all of them. A few fragmentary images can be the vehicles of thought and memory about whole scenes and events.

It is because a few details associated with a past experience can mean the whole past experience that apparently one can, in a few seconds, think of activities which took days or months to experience in the first place. All the events of a trip, or of a year in school, can appear to be remembered in an extraordinarily brief time. Actually a few details of some of the events are enough to carry the meaning of nearly all of them and to give the impression that one has remembered a great deal. Consequently, memory, imagination and dreams seem to move with tremendous speed at times.

THE CHARACTERISTICS OF IMAGERY

The processes by which images change and reproductions of past experience become fragmentary and incomplete have been subjected to a number of experimental studies. It has been found that images differ from perceptions in being generally less detailed, less stable, less clear, less intense and poorer in quality than the original perception.

Loss of detail in the image. We must note, in the first place, that images are usually limited in detail. If one calls up an image of some past experience, say the appearance of one's room on awakening in the morning, only a few of the details that were actually perceived are reproduced in the image, and some of these may be erroneous. Perhaps the appearance of the clothes chest, with one or two objects standing on it, is the only phase of the original perception that occurs in the image. In the perception one might have grasped many more details: the toilet articles on top of the chest, certain drawers left open and others closed, the mirror with reflections in it and the play of light and shadow on the wall.

Loss of stability. In addition to loss and change of detail, images are found to be less stable than perceptions. They come and go, change and are modified with great rapidity. The fleeting character of the image makes it difficult to study, for no sooner does one observe an image than it changes to a different image. The changes follow the principle of the association of ideas, one image giving way to another related to the first by similarity, contrast or contiguity. Volition, also, often plays an important part in these changes. It is easy to turn off the clatter of an imaginary alarm clock!

Loss of clarity. Images are seldom as clear as perceptions. Such details as are present in the image are usually blurred, without sharp edges and of indefinite shapes. Sometimes, in a visual image, there will be only an irregular dark gray spot in the place where we know an object to be. This

vague spot is known to represent the object but does not necessarily resemble it closely.

Loss of intensity and saturation. Even very detailed, stable and clear images are seldom as intense or, if colored, of as good saturation as the original perception. The memory image of thunder does not approach the loudness of real thunder. The memory image of the glare of the sun or an electric light bulb is seldom as intense as the original. The imaginal color of a red evening gown seems faded and washed out in comparison to the vividness of an original perception.

Perception and the image. These modifications of the image are all *changes of degree.* Compared with perception, the image suffers a loss of a quantitative sort only. Since no differences of quality exist between them, and since perceptions may be as limited in detail or quite as faint as vivid images, it is sometimes impossible to distinguish them. Particularly when stimuli are weak and fleeting, it is often found difficult in everyday life to distinguish between image and perception, and confusion occurs. For example, it may be hard to decide whether a clock is running properly, for our image of the sound of its tick may be very nearly as intense as the perception would be. At twilight or at night one may 'see' objects moving about which turn out to be apparitions. These objects are our own images which have been projected into outer space and mistaken for perceptions of real objects.

The possibility of confusion between image and perception has been demonstrated in an ingenious experiment (16). The subjects of the experiment, some of them practiced in introspection and familiar with imagery, were asked to project images of such common objects as a banana or knife on a ground-glass screen. The screen was placed in the wall between two rooms and, unknown to the subjects, a lantern slide projector was placed so as to throw pictures on the screen from behind the wall. While the subjects were trying to call up an adequate image, a very dim picture of the object was thrown

on the screen and gradually increased in illumination. Its edges were blurred, and it was made to vibrate gently. The subjects mistook this picture for an image and signaled that they had successfully projected an image on the screen. When the picture was gradually removed, the subjects continued to see an image on the screen, and this remained whether the projector was turned on or off.

The lack of distinction between perception and image frequently causes difficulty in psychological experiments. If one is testing a person's ability to hear weak tones or see small objects, images of the stimulus tone or object interfere with the experiment. It is only by rigorous control of the stimulus and repeated tests that one can tell whether the subject is actually perceiving or only imagining the stimulus.

Hallucinations. These experiences in which *images are mistaken for perceptions* are known as hallucinations. Hallucinatory images occur to everyone, but when they are frequent and occur in the absence of confusing circumstances they are symptoms of mental abnormality. Abnormal individuals suffer greatly from images of voices and sounds, visions, pressures and movements of the body. These images are regarded by such individuals as real and sometimes contribute to a disastrous dislocation of the life of the person afflicted by them.

Hallucinations and vivid imagery are common in many toxic or poisonous conditions and in states produced by drugs. The hallucinations of the alcoholic are notorious. The patient under an anesthetic has clear and intense dreams. A drug from the plant mescal produces vivid dreams or hallucinations of brilliant, saturated colors, and other drugs have similar effects (14).

Incidental memory. The degree of similarity between image and perception depends on many factors. It is clear that the factors, discussed in the preceding chapter, which influence the objective aspects of retention and memory, must also influence the subjective aspects (images) as well (see pp. 336-341). Those which merit particular emphasis in the field of

imagery are repetition of the perception, the intention to remember it and the meaning of the perception.

The first two factors are of considerable importance in practical life because many situations, which must be recalled later, *are not repeated and frequently no effort is made to remember them*. Memory under these conditions is *incidental* to other activities. Most legal testimony by witnesses of accidents and crimes is based on these incidental memories. Gross inaccuracies are apt to appear in the testimony, and this fact should be kept in mind by those involved in a legal trial (12).

Many experiments have been devised to study the accuracy of incidental memory. A typical example of such a test involves acting out a feigned quarrel or crime before a group of subjects who are unprepared for anything of the sort. A motion-picture record of the scene is compared with descriptions of it made by the subjects from memory. Obviously the trustworthiness of a subject's report depends in part upon the images that occur to him.

The accuracy of the reports varies considerably with the nature of the event. Statements about objects and persons and their activities are more accurate than statements about the characteristics of objects and persons. Statements about time and distance are found to be quite inaccurate. The most frequent errors consist in omissions of detail. The insertion or substitution of erroneous detail is slightly less frequent. Questioning the subject leads to a more complete report but also to more errors, the number of errors depending partly on the type of question asked. It is generally found that if the testimony is limited to those details of which the subject is certain, it will be more accurate than otherwise, but that errors occur even in those details on which the subject is most certain. Allowing the subject to repeat his testimony increases his feeling of certainty without increasing his accuracy. It is clear that many of the errors are due to imperfect perception resulting from emotional activity, the attitude of the perceiver, etc. Other errors arise from the meaning of the situation.

The dependence of imagery on meaning. The omission of detail and the insertion of erroneous detail in images, and in the other forms of memory, depend to a very great extent on the meaning of the situation that is recalled. This has been studied in experiments in which geometrical figures and outline drawings of objects are observed and then reproduced from memory at later periods. The results of the experiments show that much more than a simple decay of memory takes place with the passage of time. Images of the figures are *modified and changed* so that they come to resemble objects suggested by the original figure rather than the figure itself. One example of such modification has already been given in Fig. 114 of this chapter, which shows the transformation of a figure of an owl into that of a cat. The owl was apparently perceived fairly accurately, but at the time or during one of the reproduction periods its shape suggested that of a cat. The result was a modification of the memory, and the next time the owl was recalled it began to resemble a cat.

Some of the results of similar experiments (11, 3) are shown in Fig. 115. A number of drawings of relatively meaningless geometrical shapes were shown to a group of subjects. The drawings were exposed in succession for 2 seconds each. In the first column of Fig. 115 some of the stimulus drawings used in this and a related experiment are shown. To the right of each stimulus there appear a few of the reproductions of it. The reproductions of the first stimulus drawing illustrate the factor of change toward the shape of suggested objects, or *object assimilation* as it is called. This stimulus suggested different objects to different people; reproductions were given by subjects who saw the stimulus (1) as a woman's torso, (2) as a "footprint on the sands of time," (3) as a dumb-bell, (4) as a violin and (5) as a dumb-bell.

In the second and third lines of the illustration are shown the results of a related type of modification. In these the stimuli have been *analyzed verbally*, as indicated in the legend of the figure. Verbal analysis and object assimilation can be produced experimentally by saying as a figure is shown, "this fig-

ure resembles. . . ." Some of the results of such an experiment are given in the fourth and fifth lines of the figure.

Other changes than those illustrated occur. Figures are assimilated to one another, so that the reproduction contains features of each. Some features of an image are emphasized,

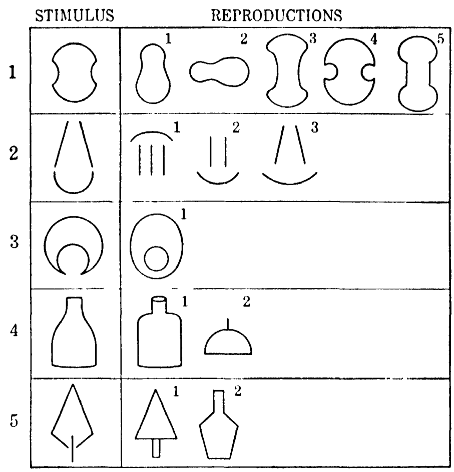


FIG. 115.

The figures in the stimulus column were shown, among a number of others, to a group of subjects. Their reproductions of these stimuli, made from memory, varied according to the meaning suggested by the figures. Thus stimulus 1 suggested (1) a woman's torso; (2) a "footprint on the sands of time", (3) a dumb-bell, (4) a violin; (5) a dumb-bell. Stimulus 2 received the names (1) pillars with curve; (2) pillars with curve (same subject as (1) but after a second exposure of the figure), (3) megaphone in a bowl. Stimulus 3 was named (1) one circle inside another. The experimenter gave different names to stimulus 4 as he showed it to different individuals. Their reproductions varied according to the name: (1) bottle, (2) stirrup. Stimulus 5 was treated as stimulus 4. The names were (1) pine tree; (2) trowel. After Gibson (11) and Carmichael, Hogan and Walter (3).

others disappear, with the passage of time. Often these changes in imagery are the results of imperfect or faulty perception. In other cases the change occurs after correct perception has been demonstrated by a good reproduction.

Most of the changes in the detail of images may be related to the influence of meaning on perception and memory. Every new stimulus is perceived in the light of the meaning the

stimulus evokes. Details which are essential to the meaning of the image are preserved and new details are even added so that the meaning is emphasized. It is probable that the meaning of a stimulus is better remembered than the percept itself. In recall the meaning serves as a cue for the particular image, and the image in turn is modified by the meaning.

THE VARIOUS FORMS OF IMAGERY

It is clear from what has been said above that images differ considerably in the accuracy and completeness with which they reproduce perceptions. These differences give a basis for classifying images into different forms. The forms of imagery, in order of similarity to perceptions, are the after-image (after-sensation), eidetic image, memory image and imagination image. The after-image is most, and the imagination image least, like the perception. These forms are closely related to one another, and no hard and fast lines can be drawn between them. It is, therefore, often difficult in practice to classify an image, especially images of memory and imagination. The after-image can, as a rule, be more sharply differentiated from the others.

The after-image. Various facts point to the *perceptual character of after-images.* They outlast the physical stimulus for a few seconds or minutes only; the other images recur long after, even years after, the perception. The similarity between perception and image is most marked in the after-image. After-images change their location with a change in the direction of regard, whereas other images are relatively unaffected by movement of the eyes. Finally, the after-image varies in size in accordance with the distance to which it is projected. This fact is known as Emmert's law, which states that the size of the after-image is directly proportional to the relation between the projection distance and the original stimulus distance. Thus if a spot of color 1 cm. sq. in size is fixated at a distance of 50 cm., its after-image will be 2 cm. sq. if it is projected 100 cm. away, 1 cm. sq. if projected 50 cm. and ½ cm. sq. if pro-

jected only 25 cm. The law breaks down if the difference between projection distance and fixation distance is too great, but holds very well for smaller differences. It is thought, because of these facts, that after-images are not true images at all, but that they result from a physiological recurrence of the stimulation after the physical stimulus has ceased, that they are, in truth, after-sensations.

Eidetic images. True images which resemble perceptions very closely and contain many vivid and clear details are known as eidetic images or primary memory images. Among adults their occurrence is rare. Children, however, have them more frequently, and investigations of imagery in children have, of late years, centered around this phenomenon.

Eidetic images, when they occur, resemble perceptions so closely that they are sometimes described as pseudo-perceptual. They are very closely related to hallucinations except that the subject knows the experience to be one of memory and does not mistake it for a perception.

Eidetic images are experimentally studied by exposing a complicated picture, such as that shown in Fig. 116, to a child. The exposure time is brief, from 10 to 40 sec. The subject of the experiment does not fixate a particular spot or keep his gaze steady, as he must to obtain an after-image, but lets his eyes roam over the picture at will. At the conclusion of the exposure the subject shifts his gaze to a gray screen and projects an image of the picture on it. He is asked to describe his image, stating what he can see in it and describing the character of the details seen.

Figure 116 was shown to 30 English school children especially selected for their clear imagery (1). Some of the children demonstrated an amazing ability to reproduce the picture or parts of it. They were able to recall in the image many obscure details in addition to those more prominent. For example, some of them were able to spell out the long German word *Gartenwirthschaft* from their image, although they did not know German. Three out of 30 children were able to spell it cor-

rectly, not only in the forward direction but backward as well. Seven more could spell it in either direction with no more than two mistakes, and they made the same mistakes in either direction of spelling. All these children could give other details of the picture as well.

FIG. 116.

This picture, used by Allport (1) in a study of eidetic imagery, was shown to a number of English school children for 35 sec. From an image of the picture the children were later able to describe accurately a very great many details, including, in some cases, the long German word over the entrance of the building.

It is this amazing fidelity of reproduction of the eidetic image that marks it out for special consideration. Yet the eidetic image, like the other forms of imagery, seldom, if ever, is a complete copy of the original. Details may be unclear or lacking. In a picture showing a boy leading a dog by a rope, one of the two figures was seen with great clarity while the other was only a blur or dark spot in the image. Furthermore, erroneous details are sometimes inserted into the image. In

these changes, interest plays a large part. Indeed an uninteresting picture may not evoke an eidetic image at all, although, for the same child, an interesting picture will give rise to a very clear image. A picture of a house, for example, failed to stimulate any image, but a picture of a monkey called out a very good one.

Eidetic images are not only clearer and richer in detail but they are also more intense and of better quality than memory images. Scientists sometimes find that some hours after having looked intently through a microscope they have a very bright image of the contents of the field. Many people report spontaneous eidetic images which occur after an evening of reading or card playing. As they go to sleep, clear and bright images of the book or playing cards are experienced.

Similarly, there are often very clear and intense auditory eidetic images. The tune that runs through one's head with great persistence for several hours or days is an eidetic image. After a conversation or argument has ceased, one hears a sentence from it very clearly and distinctly.

Memory images. The memory image is typically less clear and contains fewer details than the eidetic image. It is less stable, less intense and poorer in quality than the eidetic image. It is, therefore, less like the perception. However, memory images occur more frequently than eidetic images.

Memory and eidetic images usually possess, in addition to the characteristics already described, the characteristic of *localization*. The experience recurring in the image is localized as having originally occurred at such and such a place or time. This characteristic is usually given to the image by accompanying associated ideas. A memory of a street scene is associated with the time of year during which the town was visited. Sometimes the localization is very indirect and is accomplished only through a long chain of associations.

Many memories, however, cannot be localized. Often they are only characterized by a vague *feeling of familiarity*, which gives the knowledge that they are memories of actual previous

experiences and are not imagination images. It is thought that this feeling springs from associated motor images of the movements made in response to the situation remembered. Such motor images give a realistic and a personal quality to the memory so that it is felt as having actually happened to one's own self. The feeling of familiarity occurs in perception as well as in memory. Occasionally the feeling is experienced in situations that one knows to be new. The situation is vaguely felt to have been experienced before although one is also certain that it has not. It is probable that parts of the situation are similar or identical with situations previously encountered and that these parts call out motor imagery associated with the previous experience. The result is the feeling of familiarity in a new situation. Sometimes this quality is added to imaginary experiences, as when a person in telling a good story states it as if he had been present and repeats it until he cannot distinguish between the parts really experienced and the parts imagined.

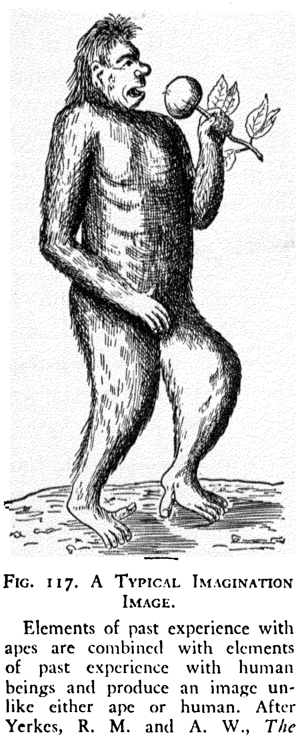


FIG. 117. A TYPICAL IMAGINATION IMAGE.

Elements of past experience with apes are combined with elements of past experience with human beings and produce an image unlike either ape or human. After Yerkes, R. M. and A. W., *The great apes*. New Haven: Yale University Press, 1932.

Imagination images. The imagination image is necessarily made up of a *combination of elements* from different memory images. Imaginary animals such as the centaur and mermaid illustrate the nature of such images. In Fig. 117 is a picture of an animal combining the features of ape and man. The early explorers of Africa were impressed by the similarity of the ape to man. In memory they

emphasized its human qualities, although some of the outstanding characteristics of the ape, such as the hairy body and long arms were retained. Parts of several experiences were put together in new relationships so that the imagination image did not resemble any single experience. It should be remembered, as was mentioned above, that nearly all memory images are also modified in some such way. A distinction can, however, be made for purposes of classification. If an image closely resembles a single previous perception it should be called a memory or eidetic image, and if it contains elements from several perceptions it must be regarded as an imagination image.

There is nothing entirely new in images of imagination except the new relationship between the parts of the image. It is impossible to imagine anything except in terms of our own past experience. Those born blind, for example, cannot imagine what it is like to see objects, except through hints derived from other senses. Imagination about life on other worlds than our own is limited by our knowledge of the forms of life and by our other experiences on this world.

Memory and imagination images make up most of the content of thought. Even concepts and abstract ideas can be represented in consciousness by memory or imagination images. The nature of *conceptual images* can be studied introspectively by thinking of a problem involving general classes of objects, such as the concept quadruped, or an abstract idea, such as the notion of center of gravity. Where is the center of gravity of a quadruped? Is it fore or aft of a line midway between head and tail? In solving the problem one has fleeting imagery representing the concept of quadruped and denoting the center of gravity. For some people the image of quadruped is a composite picture of many four-legged animals. For others, it is a memory image of a particular animal belonging to the class. Still other people invent an arbitrary construction to fit the verbal meaning or definition of the word. They put together a number of lines or blurred spots to represent head, body, legs and tail. The notion of center of gravity may also receive its imaginal representation. It will be noticed that the

images representing the concept do not differ from images of memory and imagination except in their functional value. They are individual, concrete images. Their conceptual character arises from their meaning, that is to say, the images symbolize something other than a particular experience of a particular object.

Dreams. The imagery of dreams ranges from fairly accurate reproductions of past events to the most grotesque and distorted experiences. Dreams are studied in various ways. In order to determine the *influence of external stimuli*, a sleeping person may be subjected to sounds, touches, removal of the covers and other forms of stimulation. To interpret the *meaning* of dreams, that is to say, their relation to the past experiences of the dreamer, dreams are written down immediately on awakening. They are then analyzed in relation to the life of the dreamer. A more complex method of studying the meaning of dreams is that of psychoanalysis. The individual starts from one of the events of a dream and allows his thoughts to wander freely. Each idea in the chain of associations that follow is noted down without reservation or omission. When this procedure has been completed for all the events of a dream, the trains of thought are analyzed in an effort to interpret the meaning of the dream.

In experiments to determine the effect of *stimuli* it is found that a touch on the forehead is followed by a dream of being bitten by a spider, of a headache or of being struck on the face in a quarrel. A cold stimulus applied by uncovering parts of the body evokes a dream of climbing mountains, of wading in a stream or of nakedness (17). It is also known from common observation that internal stimuli arising from cramped muscles, indigestion, worry and organic conditions frequently evoke dreaming.

It is found that dreams are related to the *recent life* of the dreamer. Dreams are most frequently drawn from the experiences of the previous day, particularly from the intense and vivid experiences. Often a train of thought concerning

past events or plans for the future is continued from the waking state into the dream state. This, also, occurs most frequently when the train of thought is vivid or characterized by emotion and excitement. Remote events are seldom directly represented in dreams, although dreams can be related to long-past events and even to some experiences that were apparently forgotten and that could not be recalled in waking life.

In addition to stimulation and past experience, the *unsatisfied desires* of an individual play a part in his dreams. Freud (9) has shown that many dreams are *wish-fulfillments.* An example of this fact is the dream of a child who, having been allowed but one piece of candy, awoke the next morning exclaiming that she had eaten the whole dishful during the night. Another example is the common dream of adults, so comforting on a cold morning, of being up, dressed and at breakfast. In these dreams unsatisfied desires are fulfilled in imagination.

Freud believes that unsatisfied desires play a part in fantastic and apparently meaningless dreams. As a result of studies by the method of psychoanalysis he has come to the conclusion that dreams in their distortion and grotesqueness symbolize the satisfaction of repressed desires. Repressed desires are those which are so unpleasant, painful or shameful that the individual tries to thrust them out of mind, hiding them from himself and others. Refusing to think about a desire does not eliminate it. The desire persists as the latent content of the dream and is able to get by what has been called the censor by appearing in distorted form as the manifest content of the dream. The dream is a disguised symbol of the desire and of the ways of satisfying it. The disguise or distortion results from the repression.

The Freudian theory holds that only the sexual desires are repressed and, consequently, that all distorted dreams symbolize sexual repressions. The term 'sex' is used in the broad sense of affection. Nevertheless the theory has been severely criticized on this point, and it is now generally believed that

any unpleasant desire may be repressed. There are other criticisms of the Freudian theory, and it is not accepted in full (4). But there is general agreement that unsatisfied desires motivate dreaming and that distortion of dream imagery occurs when these desires are repressed.

Hypnagogic images. For many people the drowsy period just preceding sleep is especially rich in imagery. Vivid, lifelike faces, landscapes, events, etc., pass before the mind. Usually these images appear without any effort to call them up, and they even occur, sometimes, despite one's will. Frequently they appear so real, perhaps as a result of their great vividness and seeming independence, that they approach or attain the character of hallucinations. They are called hypnagogic images from the name for the state between waking and sleeping.

INDIVIDUAL DIFFERENCES IN IMAGERY

The types and forms of imagery previously discussed in this chapter are not found to the same degree among all people. There are great individual differences in imagery, for some people experience images from all of the senses, whereas others are limited primarily to one or a few senses. The majority of people fall between these extremes. Most individuals, then, are *versatile* rather than specialized in their imagery (10, 7).

Likewise there are differences between individuals in the clarity and completeness of their imagery. At one extreme are a few people who have highly developed imagery from one or more senses. Thus there are many instances of apparently phenomenal accomplishments based on clear and accurate images, such as the feat of a boy who was able to give the day of the week for every date in the years 1920-1927 from eidetic images of the calendar (13). Professional memory-artists and lightning calculators often have exceptionally good visual or auditory imagery. It is probable that many people could develop similar abilities by well-directed practice.

At the other extreme are those individuals who are quite limited in the clarity and accuracy of their imagery. Many

of them claim that they think almost entirely in terms of words and the other kinds of motor images.

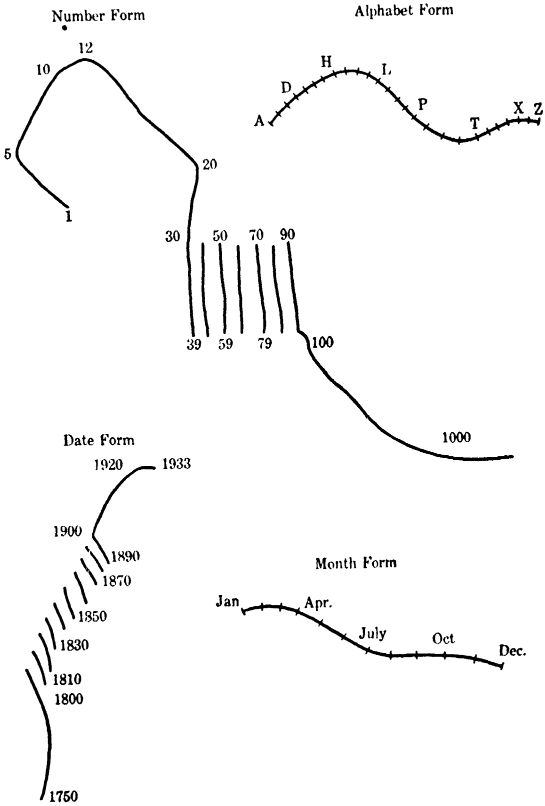


FIG. 118 EXAMPLES OF NUMBER, DATE, YEAR AND ALPHABET FORMS
The person whose number form is shown thinks of a point directly in front of his eyes when the number 1 comes to mind. 12 is further away and slightly higher than 1. 39 is back of the right elbow and the number form tapers off to the right of his body. Dates, years and the letters of the alphabet are also arranged in forms in his thinking. Courtesy of C. C. Pratt.

Number forms. The investigation of individual differences in imagery has revealed an interesting tendency in some individuals to organize series of related objects into spatial forms. Number forms are among the most common of these.

The numbers in common use are imagined as occupying a position in space. *Each number has its own position* in relation to the head, and this position is thought of when the number comes to mind. The numbers of the series are arranged in some more or less orderly and regular progression from low to high. The numbers, themselves, are not necessarily present in imaginal form, but each position in space represents or means its own number. This may be analogous to the common experience of remembering the position of words on a printed page without the words themselves actually being present in imagery.

The shape of number forms differs greatly from person to person. An example of a number form is shown in Fig. 118. It will be noticed that certain numbers, those in common use, 1, 10, 20, 30, etc., and also 12, occupy prominent positions, coming at breaks in the spatial sequence or at curves in the form. The higher numbers are usually less accurately localized than the lower, more common ones, and the number form usually tapers off to an indefinite end. Similar to number forms are month, year and alphabet forms, examples of which are also shown in Fig. 118.

Color associations. Many individuals constantly experience colors, rather than spatial locations, in association with numbers and letters. One person studied by Galton saw 1 as black; 2, yellow; 3, pale brick red; 4, brown; 5, blackish gray; 6, reddish brown; 7, green; 8, bluish; 9, reddish brown (like 6). The associations literally colored his thoughts whenever numbers were involved. The events of the century 1700-1799 were imagined as occurring on a greenish background from the color association for 7. Color associations and number forms remain quite constant from year to year. Those that experience them say that they have always experienced numbers and letters in that way.

Synesthesia. Closely related to the phenomena of color associations are those of synesthesia. Synesthesia is the tendency *to experience definite images of one sense quality when an-*

other sense is stimulated. The best-known type is colored hearing or chromesthesia. The notes of the musical scale (and other sounds) give rise to images of colors. Each note has its particular color; but notes an octave apart often stimulate the same or similar colors. In Table XIX is a report on a case of chromesthesia investigated twice with an interval of 7½ years between investigations (15). The constancy of the images, as shown in the table, is striking.

The individual whose synesthesia is summarized in Table XIX experienced a fusion of colors when two notes were sounded together. This followed the laws of color mixture (despite the fact that the individual was ignorant of these laws). The tones *g* (blue) and *a* (cold yellow) gave a hazy gray mist although at times either blue or yellow predominated. The tones *c* (red) and *f#* (blue-green) gave an indefinite light verging on gray.

Other forms of synesthesia are frequently found such as colored odors and colored tastes. A peculiar form of synesthesia was accidentally discovered during an investigation of sensitivity to cold. It was found that one of the subjects experienced

TABLE XIX

A Case of Chromesthesia Investigated in 1905 and Again in 1912

The notes of the musical scale are associated with images of very constant colors. From Langfeld (15).

	1905	1912
c....	Red	Red
d♭ ..	Purple	Lavender
d.	Violet	Violet
e♭	Soft blue	Thick blue
e . . .	Golden yellow	Sunlight
f . . .	Pink	Pink, apple blossoms
f#.	Green blue	Blue green
g♭	Greener blue	Greener blue
g	Clear blue	Clear sky blue
a	Cold yellow	Clear yellow, hard, not warm
b♭	Orange	Verges on orange
b	Very brilliant coppery	Very brilliant coppery

light, transient, pressure sensations in and around the teeth and cheeks when certain cold spots of the arm were stimulated (5). Another peculiar case of synesthesia is that of a certain psychologist, who has been blind since the age of 11. He experiences colors in conjunction with the stimuli of all the senses (19). These imaginary colors give meaning to the words and other stimuli which arouse them.

The origin of synesthetic images and of number forms is obscure, and there is, at present, no satisfactory explanation of them. Many of these constant associations of certain images and sensations are undoubtedly due to early experience. Whether experience can explain all of them or whether, in some cases, innate connections are the cause of the associations is still a matter of doubt.

REFERENCES

1. ALLPORT, G. W. Eidetic imagery. *Brit. J. Psychol.*, 1924, 15, 99-120.
2. BARTLETT, F. C. *Remembering.* Cambridge, England: University Press, 1932.
3. CARMICHAEL, L., HOGAN, H. P., and WALTER, A. A. An experimental study of the effect of language on the reproduction of visually perceived form. *J. Exper. Psychol.*, 1932, 15, 73-86.
4. CONKLIN, E. S. *Principles of abnormal psychology.* New York: Henry Holt and Co., 1927. Pp. 329-364.
5. DALLENBACH, K. M. Synaesthesis: 'pressury' cold. *Amer. J. Psychol.*, 1926, 37, 571-577.
6. DAVIS, F. C. The functional significance of imagery differences. *J. Exper. Psychol.*, 1932, 15, 630-661.
7. FERNALD, M. R. Diagnosis of mental imagery. *Psychol. Monog.*, 1912, 14, No. 58.
8. FOX, C. The conditions which arouse mental images in thought. *Brit. J. Psychol.*, 1914, 6, 420-431.
9. FREUD, S. *The interpretation of dreams* (translated from third edition by A. A. Brill). New York: Macmillan Co., 1927.
10. GALTON, F. *Inquiries into human faculty and its development.* Second edition. New York: E. P. Dutton and Co. [1907]. Pp. 57-128.
11. GIBSON, J. J. The reproduction of visually perceived forms. *J. Exper. Psychol.*, 1929, 12, 1-39.
12. HIGGINSON, G. D. *Fields of psychology.* New York: Henry Holt and Co., 1931. Pp. 465-473.

13. JAENSCH, E. R., and MEHMEL, H. Gedächtnisleistung eines schwachsinnigen Eidetikers. *Psychiat.-Neur. Woch.*, 1928, 30, 101-103.

14. KLUVER, H. Mescal visions and eidetic vision. *Amer. J. Psychol.*, 1926, 37, 502-515.

15. LANGFELD, H. S. Note on a case of chromaesthesia. *Psychol. Bull.*, 1914, 11, 113-114.

16. PERKY, C. W. An experimental study of imagination. *Amer. J. Psychol.*, 1910, 21, 422-452.

17. DE SANCTIS, S., and NEYROZ, U. (translated by H. C. Warren). Experimental investigations concerning the depth of sleep. *Psychol Rev*, 1902, 9, 254-282.

18. TITCHENER, E. B. *Lectures on the experimental psychology of the thought processes.* New York: Macmillan Co., 1909. Pp. 3-37.

19. WHEELER, R. H., and CUTSFORTH, T. D. Synaesthesia and meaning. *Amer. J. Psychol.*, 1922, 33, 361-384.

CHAPTER 15

PLEASANTNESS AND UNPLEASANTNESS

Conscious experience may be pleasant or unpleasant. Pleasantness and unpleasantness differ from sensory characteristics of experience in that they tell us but little of the nature of the outside world and much of its effect upon us. If a child is pleased by a new red toy, then both *red* and *pleasantness* are features of his experience, but it is the toy that is red and the child who is pleased. For this reason pleasantness and unpleasantness are sometimes said to be more 'subjective' than the sensory features.

Pleasantness and unpleasantness are the conscious indices of the *likes* and *dislikes* of the organism. Thus pleasantness is related to the behavior of seeking, and unpleasantness to avoidance, but this correlation is not perfect. The complexities of life present many situations in which pleasantness and its opposite do not appear to be the motivating forces.

Pleasantness is also related to beauty, and unpleasantness to ugliness; and we tend to seek beauty and to avoid ugliness. However, these esthetic values are still something else. A cathedral is beautiful but not exactly pleasant; a pain is unpleasant, not ugly. The relationship is close, but we must not identify the two terms.

HEDONIC TONE

Pleasantness and unpleasantness have toward each other a relation of opposition. If we ask an observer to arrange a group of stimuli in the order of their pleasantnesses and the stimuli vary within wide limits, we find two things: first, that the

This chapter was written by J. G. Beebe-Center of Harvard University.

pleasant stimuli are together at one end of the series, that the unpleasant stimuli are together at the other end, and that indifferent stimuli lie in between the two groups; and secondly, that, whenever it can be said of two stimuli that *f* is more pleasant than *e*, then it can also be said that *e* is more unpleasant than *f*. In other words, pleasantness and unpleasantness lie in a single linear scale with indifference at a point of juncture. Moreover, the scale must be continuous, because we can say of two pleasant stimuli that the less pleasant is also the more unpleasant. (This relationship does not hold of complementary colors. We can not say that 'less blue' means 'more yellow.')

In this way it becomes apparent that with pleasantness and unpleasantness we are really dealing with a single psychological entity, to which has been given the technical name *hedonic tone*. Hedonic tone varies in a single continuous scale from unpleasantness through indifference to pleasantness, and it is convenient to regard this scale as negative for unpleasantness and positive for pleasantness. The following table schematizes all these facts:

Stimuli	*a*	*b*	*c*	*d*	*e*	*f*	*g*
Judgments	Very unpleasant	Unpleasant	Moderately unpleasant	Indifferent	Moderately pleasant	Pleasant	Very pleasant
Hedonic tone . . .	−3	−2	−1	0	+1	+2	+3

In experimental research it is found that observers can quite accurately make judgments of hedonic tone on a scale that varies from —3 to +3 in accordance with the implications of the foregoing table. It would be ideal, of course, to build up a much more elaborate scale of measurement in which the successive steps represent only just noticeable differences of pleasantness and unpleasantness, but such a procedure is too laborious to be practical.

Psychologically, hedonic tone is to be understood by a study of the conditions which determine it. In part these conditions are 'internal,' that is to say, factors of learning or motivation are operating. We shall consider such conditions later in the

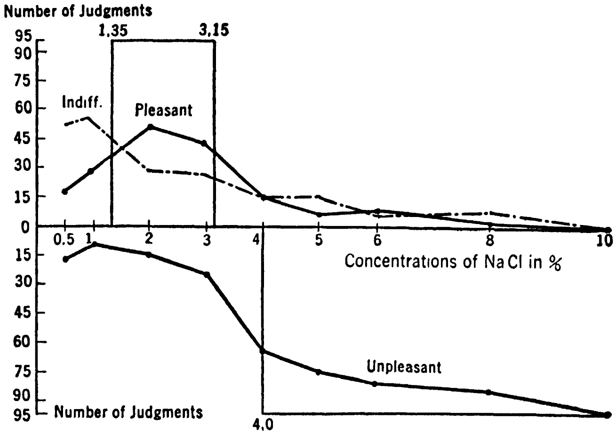


FIG. 119 HEDONIC TONE OF COMMON SALT (SODIUM CHLORIDE) SOLUTIONS AS A FUNCTION OF STRENGTH OF STIMULUS

Seven observers in 7 sittings. The figure shows also the ideal stimulus-ranges for pleasantness and unpleasantness. After Engel

present chapter. On the other hand, there is a great deal for us to learn about the relation of hedonic tone to external stimulation, a topic which has yielded with comparative ease to experimental investigation. Here, for convenience, we must distinguish between primary and secondary stimuli. The *primary stimulus* is that part or aspect of the total stimulating environment to which conscious attention is given and which seems to the observer to be the crucial stimulus. The *secondary stimulus* is all the rest of the effective stimulation. When a green square is said to be the hedonic stimulus, it is usually only the primary stimulus, the immediate object of attention to the observer. The secondary stimuli include the visible surroundings of the square, the preceding visual stimulation, the accompanying auditory or tactual stimuli, etc. We shall deal first with the relation of hedonic tone to primary stimuli.

HEDONIC SUMMATION

There is hedonic summation. When two simple stimuli are combined in some total perception, the hedonic tone of the

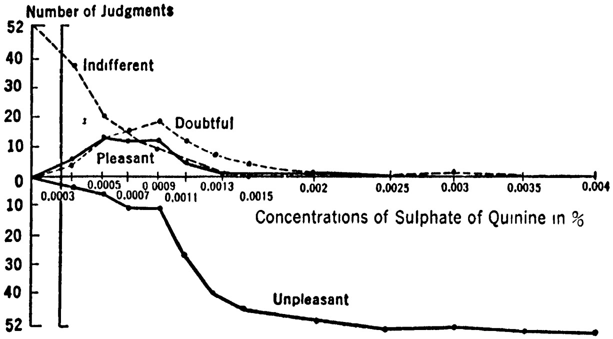


FIG. 120. HEDONIC TONE OF SOLUTIONS OF SULPHATE OF QUININE AS A FUNCTION OF STRENGTH OF STIMULUS.

Ten observers in 6 sittings. After Engel

resultant depends upon the sum of the hedonic tones of the two stimuli. We must, however, limit this principle in two ways. (*a*) It holds only when it can operate without interference by some of the other hedonic laws explained in this chapter. (See especially the effects of hedonic contrast, pp. 383f.) (*b*) It holds only when the relative importances of the different constituents in the sum are determined and taken into account. Thus the correct formula is

$$H \sim w_1 h_1 + w_2 h_2,$$

where H is the resultant hedonic tone; h_1 and h_2 are the hedonic tones of the constituents, and w_1 and u_2 are 'weights' which determine the relative effects of the two stimuli. Thus one investigator of the hedonic tone of odors found that his results could be accurately expressed by the equation $H = 0.54\,P + 0.69\,U - 0.21$, where P and U stand for the hedonic tones of pleasant and unpleasant components. Formulas of this sort do not have great predictive value, because the weights have to be independently determined for every new set of stimuli, but the fact that they can be validated means that the concurrent operation of hedonic stimuli involves a relatively simple principle.

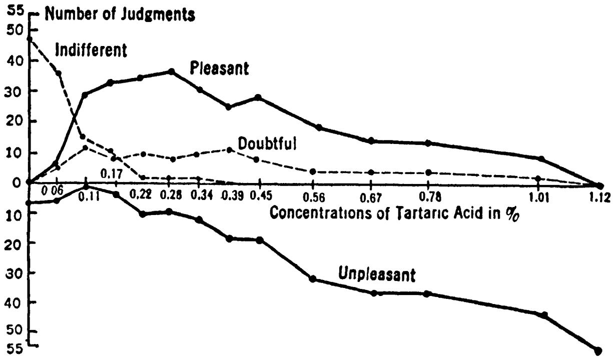


FIG. 121 HEDONIC TONE OF SOLUTIONS OF TARTARIC ACID AS A FUNCTION OF STRENGTH OF STIMULUS.

Seven observers in 6 sittings. After Engel.

This principle of summation does not hold for the rapid succession of colors or for concordant tones. In the case of the tones we soon become involved in the complexities of musical harmony. There are no such intricacies in the simultaneous combination of colors and odors, or of a color with a shape. The beauty of a dress is not really so complexly determined as to entitle it to be called 'a symphony of color.'

INTENSITY AS A CONDITION OF HEDONIC TONE

Hedonic tone tends to vary with the intensity of the stimulus. As stimulus intensity increases, hedonic tone increases to a maximum of pleasantness, and then decreases, passing through indifference to great unpleasantness for very intense stimuli. Figs. 119-122 show this function for stimuli to the four primary qualities of taste. That cane sugar (Fig. 122) does not become unpleasant in the higher concentrations may be due to its limitations as a stimulus, since cane sugar in maximal concentrations is not so sweet as some other substances. The curves make it clear that the rule holds on the average: sometimes a weak stimulus is unpleasant, although the general trend is unmistakable.

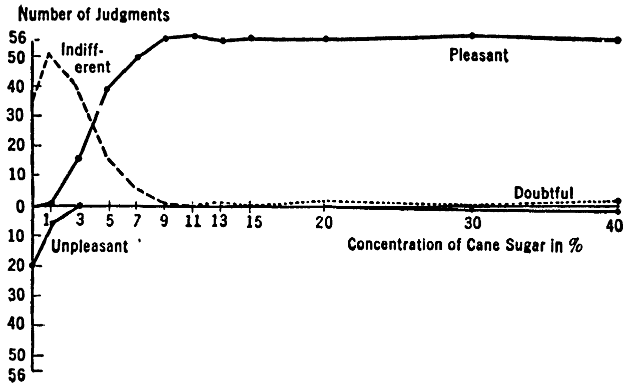


FIG. 122. HEDONIC TONE OF SOLUTIONS OF CANE SUGAR AS A FUNCTION OF STRENGTH OF STIMULUS

Ten observers in 6 sittings. The function might pass over into unpleasantness if it were possible to get a more intense sweet from cane sugar. After Engel

According to the data of Figs. 119-122 the most pleasant concentrations of the four sapid substances are:

Cane sugar	9.	per cent
Common salt	2.	per cent
Tartaric acid	0 28	per cent
Sulphate of quinine.........	0 0007	per cent

The biological use of this general rule for intensity is obvious. Intense stimuli tend to be dangerous, and unpleasantness may act as a warning. Pleasantness reflects the happy mean between no stimulation at all and extreme, violent, dangerous stimulation.

QUALITY AS A CONDITION OF HEDONIC TONE

Although it is plain that in particular cases the quality of the stimulus is the important condition of its hedonic tone, exact information on the matter is lacking because most of the experiments have failed to take account of intensity. As we have just seen, all qualities tend to change in hedonic tone as intensity changes. However, certain generalizations can be made for the five departments of sense.

Color. Saturated colors are usually preferred to unsaturated ones. Red and blue are usually the most pleasant hues, and yellow the least pleasant. When colors of similar brightness and saturation are combined in pairs, hues that are nearly complementary are most often preferred. However, when the pair differs greatly in saturation or brightness, small differences of hue tend to be preferred. When gray is paired with a color, red and blue continue to have the advantage. Of course these gross generalities are constantly being cut across by other laws. It would not be maximally pleasant to have every room in every house decorated in red or blue.

Tone. Single tones show but little hedonic differentiation, but tonal combinations differ greatly, as the facts of music would lead us to expect. The following table shows the average order of preference among a large number of observers for the twelve musical intervals that lie within the octave. The table begins with the most pleasant and ends with the least pleasant.

1 Major third	4 Major sixth.	7. Tritone.	10 Minor seventh.
2 Minor third.	5 Minor sixth	8. Fifth	11 Major seventh
3. Octave.	6 Fourth.	9 Major second.	12 Minor second.

Taste. In the preceding section we have seen how hedonic tone is a function of gustatory intensity. These same facts (see Figs. 119-122) seem to imply that in ordinary concentrations sugar is the most pleasant of these four stimuli and quinine the least pleasant. In fact, it is in the dictum of common sense that sweets are pleasant and bitters unpleasant, although we now know that weak bitters, like olives and coffee, may be pleasant.

Touch. Pain is notoriously unpleasant, although a very weak pain, like the prick of a carbonated drink, may be pleasant. Roughness is apt to be unpleasant and smoothness pleasant. Mild pleasantness was induced in one experiment by stroking the observer's forehead with velvet.

Smell. In general, odors that belong in or near the fruity, flowery, spicy and resinous classes are pleasant and the odors in the putrid and burned classes are unpleasant. In other words,

one face of the smell prism (Fig. 46, p. 147) tends to be pleasant and the remaining edge unpleasant. On the other hand, some people like onions, which belong nearest the putrid class, and not everyone likes flowery-fruity geranium or spicy cinnamon.

FORM AS A CONDITION OF HEDONIC TONE

Experimental esthetics and the practice of artists have established numerous facts concerning the hedonic tone of forms. The most important principles in the case of temporal forms are the rules of harmony and melody formulated by musicians. One important fact that has to do with spatial form is the esthetic superiority of the 'golden section,' *i.e.,* the division of a line at such a point that the smaller part bears the same ratio to the larger part that the larger part bears to the whole. However, most attempts to generalize from these principles have not led to the formulation of rigorous laws.

There is an old theory in esthetics that beauty is dependent upon 'unity-in-multiplicity.' Unification tells for beauty. Complexity alone precludes monotony but it also tends toward chaos, so that it is only in the presence of unity that it yields beauty. Recently a mathematician (Birkhoff, 4) has sought to give quantitative exactness to this vague theory. He has laid down the rule that the esthetic measure of an object, M, is proportional to its orderliness, O, and inversely proportional to its complexity, C:

$$M = \frac{O}{C}.$$

He then constructed quantitative definitions of O and C, which he applied separately to melodies, drawings of vases and drawings of polygons. The test of the adequacy of his definitions and of the general principle for which they were created lies solely in experimental verification. For us the question is: does hedonic tone consistently vary with the value of M? The test has already been made for the polygons, with

the result that it appears that the computed *M* accords as closely with the average judgment of a group of observers as does the judgment of a single observer agree with the average of the group. In other words, for complex forms unity-in-multiplicity is a very important condition of pleasantness.

HEDONIC INDIVIDUAL DIFFERENCES

It is common knowledge that persons differ greatly in their 'tastes.' All sorts of individual, 'internal' factors, which depend upon the motivational systems of the personality and have been learned or matured in the life of the individual, become effective and account for the variability in the laws which we have just considered. That the cause of this variability may be largely personal is shown by the fact that persons tend to differ more from one another than does one person change in his hedonic judgments from time to time. For instance, in one experiment eight observers arranged 14 colors in rank-order on each of two occasions. The coefficients of correlation (see pp. 45f.) between the orders of pairs of observers ranged from 0.88 to —0.21, that is to say, there was, on the average, hardly any significant relationship at all. But the coefficients of correlation between the orders for the same observer at two different times ranged from 0.69 to 0.96.

DEPENDENCE OF HEDONIC TONE UPON SECONDARY STIMULI

The primary stimulus—the stimulus in the focus of the observer's attention—is not the only stimulus to influence hedonic tone. When an observer attends to one of a pair of colors, the hedonic tone of the pair is influenced principally by the color to which he attends, but the other color also has some effect. Moreover, there is evidence that the hedonic tone of a stimulus may be altered by the simultaneous stimulation of even a different department of sense. However, the most striking evidence of the effect of secondary stimulation is to be found in

the way that the hedonic tone of a present stimulus depends upon past stimulation.

This is the principle of *hedonic contrast.* The pleasantness of a given stimulus depends upon the hedonic tone of the preceding stimuli. Let a given stimulus follow one group of stimuli on one occasion, and a more pleasant group on another occasion, then the given stimulus will be less pleasant on the second occasion. However, the rule holds only when the given

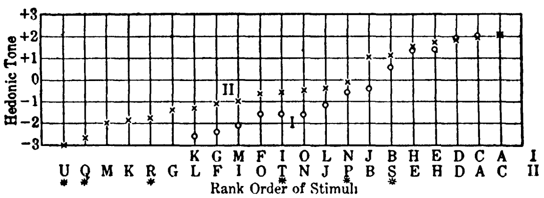


FIG. 123. HEDONIC CONTRAST.

Sample data for a single observer. Hedonic scale (ordinates), — 3 to + 3. I: Circles show hedonic rank-order of 15 stimuli, *A . . . O*, in order shown in scale I. II: Crosses show rank-order of same 15 stimuli, and 6 additional stimuli, *P . . . U*. The order of the entire 21 is shown in scale II, where the additional stimuli are marked with asterisks. The rank-orders of stimuli *A . . . O* are not greatly changed from the one experiment to the other (compare scale I with II), but the hedonic tone of the original 15 stimuli, *A . . . O*, is raised by the addition of the 6 less pleasant stimuli, *P . . . U*. While II would seem to have greater hedonic range than I, both are scattered uniformly between — 3 and + 3, thus illustrating the fact that hedonic judgments are made relativistically in respect of the entire group of stimuli. The data and functions are from the work of N. E. Cohen.

stimulus is perceived as belonging to the group of stimuli that affects its hedonic tone. There can be this sort of hedonic contrast among colors or among odors, but not, it seems, from one sense-department to another. On the other hand, when the nature of a unitary group is established, the mutual effects of the stimuli upon one another persist over relatively long intervals of time. For instance, a color which in many situations would be only moderately pleasant can be made very pleasant by grouping it with a number of less pleasant colors, and this color will then again be perceived as very pleasant, even as much as two days later, provided at that time it is recognized as belonging to the group of less pleasant colors.

As a matter of fact, hedonic contrast turns out to be a special case of a more general psychological principle. In order to understand this relation let us consider the experiment for which Fig. 123 gives a sample set of results. In the case in question, 15 stimuli had been presented to an observer and he had given judgments of hedonic tone for them, so that the experimenter was able to arrange them in a rank-order of pleasantness. These stimuli are those designated by the letters, *A, B, C, D, . . . O* in scale I of the figure, and they were, on the whole, stimuli that would usually be pleasant. The function represented by circles in Fig. 123 shows the rank-orders that these 15 stimuli received in respect of hedonic tone, and it is seen that the function approximates a straight line running from —3 to +3 in the hedonic scale. After this experiment 6 more stimuli, *P, Q, R, S, T* and *U*, all normally unpleasant, were added to the original 15, and the rank-order for the entire 21 was determined in a similar manner. Scale II and the function represented by crosses show the result. What happened?

What happened was, as the figure shows, that the addition of the less pleasant stimuli, *P . . . U,* forced the more pleasant stimuli up in hedonic tone, in such a way that both the smaller and the larger group were approximately uniformly distributed between —3 and +3 in the hedonic scale. There was contrast: the unpleasant stimuli enhanced the pleasantness of the more pleasant stimuli. But the apparent contrast is really only an instance of the relativity of judgment. These judgments that appear to the observer to be 'absolute' are really relative to the total group of stimuli. The least pleasant stimulus of any group tends to be regarded as unpleasant and reported as —3. If still more unpleasant stimuli are added, then the hedonic status of the previously least pleasant stimulus is improved.

After all, this relativistic principle is familiar to everyone. There is a man, let us say, who finds the food at his boarding house very unpleasant: it is the worst food that he has to eat. Then he lands in jail and the food is, as food, still worse. Yet it soon becomes only the worst food that he has to eat.

When he goes back later to the boarding house, the food seems much better as long as he does not forget about the food at jail—that is to say, in our technical terms, the boarding-house food is better as long as it continues to constitute with the jail food a unitary group. But if the man forgets all about jail, then the table at the boarding house is likely to seem to deteriorate again.

The same relationships work out at the other end of the scale. Our greatest pleasures fade before still greater ones, and the still greater ones do not remain 'still greater,' but become simply the maximum of which the particular person is capable, the '+3' on the scale. Here lies one reason why wealth cannot assure happiness.

It is probably in phenomena of this sort that we have the ground for the misleading belief in hedonic habituation. It is commonly supposed that very pleasant or very unpleasant objects approach indifference on being repeatedly experienced. However, there is no sufficient experimental evidence to support such a view. Repeated gustatory stimulation with children has shown an approach toward hedonic indifference, but the repetition of classical musical compositions tends to enhance enjoyment. It is possible that the belief in hedonic habituation has come about because the extreme pleasures and pains of anticipation become a mere '+3' or '—3' in realization according to the general relativistic principle.

Relativity of judgment is not a phenomenon of pleasantness and unpleasantness only. It seems to apply to all judgments. The intensities of sounds and of weights depend in the same way upon the other intensities that form with them a unitary psychological group.

It is probable that relativity of this sort is a special case of the adjustment of the organism to its environment. The organism minimizes the frequency of extremes by getting away from them or by adjusting to all but the most extreme. This adjustment is a kind of hedonic equilibrium. Rich man and poor man, saint and sinner, all find their pleasures and dis-

pleasures spread out from '—3' to '+3.' It is the business of the organism to maintain a balance of feeling.

DEPENDENCE OF HEDONIC TONE UPON MOTIVATION

We have already noted that hedonic tone depends, not only upon both primary and secondary stimulation, but also upon 'internal' factors which represent the state in which stimulation finds the organism. In turning our attention now to such conditions of hedonic tone we shall consider first three general cases of what may be called motivating factors.

Cognitive attitude. The point of view or cognitive attitude of the observer at the time of stimulation may affect hedonic tone. A good many different attitudes are differently effective in this way, but we need to consider here only four whose consequences were definitely established in a certain experiment.

There is, first, the *objective* attitude, in which observation is directed to the sensory aspects of the stimulus, as, for example, the hue, brightness and saturation of a color. This is the proper attitude for the observer to take when the effect of a primary stimulus is under experimental investigation, but it elicits less extreme hedonic tones than do the other three attitudes.

Then there is the *physiological* attitude, in which observation is directed, not to the stimulus, but to the effects of the stimulus upon the person of the observer, as, for example, the production of goose-flesh or nausea or excited breathing.

The *associative* attitude is very common and consists of attention to the verbal or imaginal associations that may, when permitted, crowd in upon the primary experience. A green stimulus-paper is not likely to have the same hedonic tone if it means *jade* to the observer as it has if it means *bile*.

Finally there is the *character* attitude, in which the observer reads into the stimulus personal characteristics, like cheerfulness or depression. A particular yellow may be the least pleasant of the hues, but the same yellow could be quite pleasant when perceived as a 'cheerful yellow.'

Each of these four attitudes is likely to give a different set of hedonic tones for the same set of stimuli. Let us imagine them all applied to the same stimulus, a green paper, by the same observer, a young woman. She sees the stimulus as a dark, well-saturated, slightly smudged green (objective attitude) and reports slight unpleasantness, —1, let us say. Another time she sees it as a vile, repulsive color, a little sickening perhaps (physiological attitude), and she reports it as very unpleasant, —3. At still another time the color has no bodily effect upon her, but it appears to be a despondent, depressing color (character attitude), and she judges it to be unpleasant, —2. On a final occasion she perceives it as the exact color of the velvet dress which she had admired so much the evening before (associative attitude), and now it seems pleasant to her, +2.

Physiological balance. The disturbance of any important physiological balance of the organism acts as a need, and a need like other motivational factors affects hedonic tone. If the organism is deprived of water it becomes thirsty, which is unpleasant; and, if it quenches its thirst, that is pleasant. The perception of water has an hedonic value for the hot and thirsty traveler different from that which it has for the man who lives beside the well.

Hunger and thirst are the best-known needs that arise from lack of physiological balance. The need for salt is even more specific though not so often realized by man. And there are many other physiological needs.

The operation of these needs upon hedonic tone seems to occur indirectly by way of sensory mechanisms. The need of the body for water is represented by an unpleasant, dry, burning pressure in the mouth, but the sensation and the hedonic tone both disappear as soon as water is placed in the mouth, even though the need of the body is as yet unsupplied. Fasting at first brings on hunger contractions of the stomach and an unpleasant experience of hunger, but after a while the sensation and the unpleasantness go away, although the need of the

body for food remains. (See p. 186 for the sensory nature of hunger and thirst.)

Volitional attitude. It is plain that hedonic tone is intimately connected with success and failure, and thus with what we may call volitional attitudes—for failure is a thwarting of the will, and success is an achievement of the will. In certain well-known experiments the psychologist Ach (1) formulated laws of the relation of the will to hedonic tone, and we may here reformulate the substance of these rules in the familiar terms of our common everyday experience. The laws are as follows. (*a*) Success is pleasant. (*b*) Failure is unpleasant. (*c*) The stronger the will to succeed, the more pleasant is a success, or the more unpleasant is a failure. (*d*) The stronger the opposition to success, the more pleasant is a success. (*e*) The greater the success, the greater is the pleasantness. Ach's experiment simply puts the seal of validity upon principles that are well recognized in everyday life.

DEPENDENCE OF HEDONIC TONE UPON MATURATION AND LEARNING

Everyone knows that 'tastes' change as persons mature, and it is easy to see that many such changes must occur as the result of the development of the motivational factors which we discussed in the last section. On the other hand, some changes, like those for color preferences, seem to be more directly dependent upon the stimulus.

In the case of *colors*, preferences during the first half year of life appear to depend entirely upon illumination: the infant grasps at the brighter of two objects irrespective of their hues and saturations. During the second half year the infant develops a preference for the warm colors, red and yellow, over the cold ones, green and blue. In the second year this distinction very gradually begins to disappear, so that by adolescence the preferences of children are like those of adults, for whom red and blue are preferred to yellow and green.

There are tests which show that preferences for *odors* do not change significantly with increasing age after the eighth year

of life. Unfortunately these tests do not extend back into early childhood and infancy.

Preferences for *musical intervals* have been found not to change significantly as a function of age after the seventh year, but there probably is some shift with musical education. Musical persons definitely prefer the thirds and the sixths to the fourth and the fifth, whereas unmusical persons sometimes do not. On the other hand, children under ten years old without musical training often like discords.

It has been shown for *taste* that stimulation with sweet has a calming effect upon a new-born child, that salt results in slight unrest, and that sour and bitter result in marked upset. These effects persist essentially throughout life.

It is difficult to say how much such changes depend upon the natural maturation of the organism and how much upon specialized learning. In any case we must not overlook the fact that *learning* plays a very important role in the establishment of the conditions of hedonic tone. 'Taste' can be learned—as the estheticists have always held, and as the changing hedonic value of women's dress to the current fashion repeatedly attests.

A special case of the dependence of hedonic tone on learning is to be seen in the 'conditioning' of emotions. A classical example in the experimental literature shows how an infant, accustomed to play happily with a white rat, learned to be afraid of the rat and to greet it with crying and vigorous movements of avoidance. However, the gross fact, that emotions depend upon past experience and learning, does not require experimental demonstration. Abundant evidence is to be found in the courts, the marriage clinics, the offices of diplomats and politicians, the offices of psychiatrists and every place where human emotions arise unexpectedly to defeat wisdom.

RELATION OF HEDONIC TONE TO RESPONSE

There is a belief of long standing among psychologists that hedonic tone gives rise to, or at least is associated with, certain responses—muscular movements or glandular secretions. It

was hoped that these responses could be regarded as 'expressive movements,' symptomatic indices of hedonic tone. Most of these hopes have been disappointed, but there remains still a core of plausible fact with which this section is concerned.

Vasomotor and respiratory changes. The net outcome of a large body of conflicting researches on the relation of vasomotor and respiratory changes to hedonic tone is shown in Table XX. Even here, some queries have to be indicated, and the other statements of the table represent only the preponderance of the evidence, since every item has been questioned at some time or other.

TABLE XX

VASOMOTOR AND RESPIRATORY CHANGES ACCOMPANYING PLEASANT AND UNPLEASANT STIMULATION

	Pleasantness	Unpleasantness
Vasomotor changes		
Rate of pulse	decreased	increased
Height of pulse	increased	decreased
Volume of arm*	increased (?)	decreased
Respiratory changes		
Rate of breathing.............	increased	decreased (?)
Depth of breathing...........	decreased (?)	increased

* The volume of the forearm changes as the result of changes in blood supply due to vasodilation or vasoconstriction. It is easily measured by a special instrument, the plethysmograph, which also records the pulse

Galvanic skin response. The galvanic skin response, the change in the electrical resistance of the skin, often accompanies pleasant and unpleasant stimulation, and has therefore been thought to be a good indicator of hedonic tone. Although it may be true that the occurrence of this response establishes a presumption of the existence of hedonic tone, still it is not a good indicator. It does not differentiate between pleasantness and unpleasantness. Moreover, its absence does not create a presumption of the absence of hedonic tone, for pleasantness and unpleasantness may occur without the re-

sponse. Thus it seems probable that this relationship is quite indirect, and that the galvanic skin response is more closely associated with innervation of the sympathetic nervous system, which is generally involved in emotion and is itself related in excitation to pleasantness and unpleasantness (see pp. 411-415, 416f.).

Locomotor response. There are no known *simple* correlations of locomotor response with hedonic tone. Certain classical views have been disproved. A man is not necessarily stronger when pleased than when displeased; nor does he invariably make movements of extension in pleasure and of retraction in displeasure. However, there is one simple rule. When he likes a thing, man tries to get it if he can. When he dislikes a thing very much, he tries to get away from it, whether he can or not.

The grounds for this rule lie in the results of an experiment. An observer was shown hedonic stimuli under different conditions. When the conditions were such as to render movement of the observer impracticable, a pleasant stimulus led to muscular relaxation and an unpleasant one to muscular tension. These cases seem to correspond to passive enjoyment for a pleasant object and to restlessness and strain for an unpleasant object in situations where activity is impossible or useless. On the other hand, when the observer was left free to move and the stimulus was moving past him, then a pleasant stimulus elicited actual approach and an unpleasant stimulus actual withdrawal. Here we have the more usual case where pleasantness and unpleasantness are associated respectively with seeking and avoidance.

It is plain that this entire matter is tied up with the problems raised in *hedonic theories of action.* Such theories assume that hedonic tone determines action. Some theories hold that present action depends upon the part played by the hedonic tones which were present when action patterns were being formed; others hold that present action is directly in accordance with present hedonic tone; still others maintain that

action is determined in respect of hedonic tone anticipated in the future. However, all the theories depend on the correlation of pleasantness with seeking, approach, acquisition; and the correlation of unpleasantness with avoidance, withdrawal, rejection. There can be little doubt that we are facing here a fundamental fact, although it is complicated and limited in innumerable ways in adult human life.

DEPENDENCE OF LEARNING UPON HEDONIC TONE

Hedonic tone, as we have seen, depends upon learning, but there is also evidence that learning depends, at least in part, upon hedonic tone. Pleasure seems to stamp learning in, and consequently pleasant events are apt to be best remembered. While the psychology of learning is indeed a very much larger topic than the relation of hedonic tone to memory (see chap. 13), it is interesting to note that this mutual relation between learning and pleasantness explains one way in which education takes place. The child who likes candy learns to like his aunt who gives him candy: hedonic tone depends on learning. So the child learns the multiplication table because his aunt, whom he now likes, teaches it to him and gives him the pleasure of her approval for a reward: learning depends on hedonic tone. By such transfers of hedonic values it might be possible to make a mere lover of candy into a mathematician, provided certain other capacities were available.

There is considerable evidence from the psychological laboratory that pleasantness is conducive to good memory. The recall and recognition of previously presented objects are better for pleasant than for unpleasant objects. Pleasant poetry is learned more readily than unpleasant. The association experiment shows that associations come more rapidly for pleasant than for unpleasant words. Persons who are asked to recall the events for some particular period of time, like a summer vacation, report more pleasant than unpleasant items. It may be, however, that this last fact is not significant, for one psychologist, after studying special diary accounts, came to the

cheerful conclusion that there are more happy than unhappy events in life.

Not only does pleasantness reinforce learning, but unpleasantness hinders learning as well. For this reason either *reward* or *punishment* can be used in the establishment of learning.

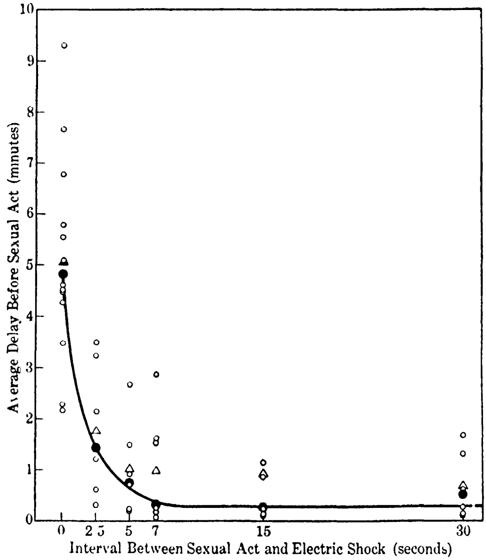


FIG. 124. EFFECT OF IMMEDIACY OF PUNISHMENT UPON LEARNING.

The abscissae represent for a training period the lapse of time between the initiation of sexual activity in a male rat and the giving of an electric shock to the rat, *i.e.*, the delay of punishment. The ordinates represent the subsequent delay of the rat in initiating the sexual act. The circles represent individual cases, the triangles means and the dots medians. The black curve is based on the medians. In general, the rat may delay for 5 min. when the shock has followed activity at once. The delay becomes less as the interval becomes longer, and reaches the normal time of about 15 sec. at an interval of about 7 sec. After D. R. Wheeler.

Experiments upon animals show that the task that is followed by the greater reward or by the less punishment is the task that is learned. Moreover, the reward or punishment is more effective the more immediate it is.

Fig. 124 shows how important immediacy is. The curve describes the results of an experiment in which the normal sexual behavior of male rats was inhibited by an electric shock. In a

training period the rat was given the shock after the sexual act at intervals that varied from 0 to 30 sec. (horizontal scale of Fig. 124). After this training it was observed in test trials how long the rat would delay sexual activity when the opportunity was provided. The delays varied from a few seconds up to nearly 10 min. (vertical scale of Fig. 124). The general trend of the function is plotted in the figure and leads one to say that 7 sec. is the limit of immediacy. If the shock follows the act by less than 7 sec., the rat takes the unpleasantness as if it were a consequence of the act and delays before acting again. And, the greater the immediacy, the greater the delay.

PHYSIOLOGY OF HEDONIC TONE

It is now a generally accepted physiological theory that hedonic tone has something to do with the region of the thalamus in the brain. There is a great deal of evidence, but it is not conclusive as to detail.

In 1887 Bechterew observed that a dog from which the cerebral hemispheres had been removed could still exhibit the behavior which is usually taken as indicative of pleasantness and unpleasantness. He also found that this hedonic behavior disappears when the thalamus is destroyed. Later experimentation has confirmed his finding, although Bard (2), as the results of experiments with rage, has concluded that emotional behavior depends on the preservation of the hypothalamus rather than the thalamus.

Further evidence comes from the study of patients who have developed brain lesions which are later accurately localized in the brain. Lesions of the cortex do not lead to hedonic abnormalities, but lesions of the thalamus do. Head and Holmes (5), pioneers in this field, have described many cases of exaggerated hedonic susceptibility as a result of thalamic lesion For instance, they reported for cases affected on only one side that "many of these patients complained that they could not be shaved on the affected cheek because it seemed as if the razor were passing over a raw surface." Nor was this exag-

geration confined to unpleasantness. Warmth frequently elicited greater pleasantness on the affected side than on the normal side. "In one case, a tube containing water at 38°C applied to the normal palm was said to be warm, but the patient cried out with pleasure when it was placed in the affected hand."

Beyond this gross statement that the functioning of the thalamic region is very important for hedonic tone it is not safe at present to go. The more detailed theories of the relation of the thalamus to the cortex, when pleasantness and unpleasantness are experienced, lack confirmation.

There is some evidence that bears on the sensory nature of hedonic tone. Nafe (6) found experimentally that pleasantness can be described as a "bright pressure" and unpleasantness as a "dull pressure." His results have been both contradicted and confirmed. There is a possibility that the bright and dull pressures merely accompany the hedonic tones but are not to be identified with them. Whatever the exact truth turns out to be in this controversy, the research has served to emphasize the fact that hedonic tone is very much like somesthetic sensation, or is perhaps a kind of somesthetic sensation. It is this conclusion that gives ground for hesitation in assuming that the thalamus or hypothalamus acts entirely independently of the cortex when pleasantness or unpleasantness is felt.

There can be little doubt that eventually all the facts of hedonic tone that are mentioned in this chapter will be brought together in some physiological theory. At the present time, however, any such attempt would lead to unwarranted speculation, and we shall do well to rest content with the ideas that hedonic tone depends upon the functioning of the thalamic region and that it is sensory in its general character.

REFERENCES

The text supplementing most directly and explicitly the subject-matter of the present chapter is Beebe-Center (3).

1. ACH, N. *Ueber den Willensakt und das Temperament.* Leipzig, 1910.
2. BARD, P. The neuro-humoral basis of emotional reactions. In *Handbook*

of general experimental psychology. Worcester, Mass.: Clark University Press, 1934. Pp. 264-311.
3. Beebe-Center, J. G. *The psychology of pleasantness and unpleasantness.* New York: D. Van Nostrand Co., 1932.
4. Birkhoff, G. D. *Aesthetic measure.* Cambridge, Mass.: Harvard University Press, 1933.
5. Head, H., and Holmes, G. Sensory disturbances from cerebral lesions. *Brain,* 1911, 34, 102-254.
6. Nafe, J. P. An experimental study of the affective qualities, *Amer. J. Psychol.,* 1924, 35, 507-544.
7. Troland, L. T. *The fundamentals of human motivation.* New York: D. Van Nostrand Co., 1928.
8. Troland, L. T. *The principles of psychophysiology.* New York: D. Van Nostrand Co., 1932. Vol. III.
9. Valentine, C. W. *An introduction to the experimental psychology of beauty.* Revised edition. London, 1919.

Chapter 16

EMOTION

Scientific knowledge concerning emotion is neither comprehensive nor overly exact. Human interest in emotion, on the other hand, is universal. This disparity between universal interest and meager psychological knowledge is a continuous challenge to the scientist.

THE NATURE OF EMOTION

Since emotion is a word or a name there is always the tendency to think that emotion is some *thing*, some discrete, distinct, definable unity. Experiments have uniformly shown that emotion, like perception, cognition and attention, does not exist as a unique entity, but that we find a variety of emotional experience and reactions, just as we find a variety of perceptive experience, attentive reactions and cognitive experience. It is also true that such emotions as anger, fear, pity or disgust do not exist in unique independence, but that the phenomena that we find are such things as fighting reactions or fearfulness, the experience of pitying or the withdrawal from obnoxious objects. *Neither emotion nor emotions exist as discrete entities.*

The common characteristics or criteria of emotion are not simple or straightforward. Primarily one thinks that marked or intense pleasantness or unpleasantness (approaching or withdrawing reactions) pervades emotion, but exceptions occur. In such a situation as an intense hand-to-hand combat the participant usually experiences neither pleasantness nor unpleasantness during the fight. Emotional reactions are usually

This chapter was written by Carney Landis of the N. Y. Psychiatric Institute and Columbia University.

not specific but involve practically the entire organism. Again there are exceptions. The cautious, well-planned reactions which occur in fearful escapes are narrowly directed upon a particular end. Emotional experiences or reactions are usually accompanied by some involvement of visceral or organic change, but occasional cases occur in which there is a total motor and sensory paralysis below the shoulders, after injury to the spinal cord, with retention of emotional experience and expression. It is true that various theorists have attempted to locate a distinct, unique, experiential element in one or another specific portion of the nervous system or in some particular relationship between nervous impulses, but thus far these hypotheses have not received experimental verification or else are experimentally unverifiable.

It seems that emotion can best be characterized as a *relationship existing between many diverse elements of experience and reaction.* This relationship is not well specified, but, generally speaking, there is marked pleasantness or unpleasantness and disorganization of usually integrated behavior patterns. An emotion is the total of the experience of an individual during any period of time when marked bodily changes of feeling, surprise or upset occur.

THE DESCRIPTION OF EMOTIONAL BEHAVIOR AND EXPERIENCE

In spite of the great number of literary and dramatic descriptions of emotion that are available, it is not easy to secure material which adequately and accurately pictures emotion. This difficulty is due to the fact that the selection is out of context. *For an emotion to be an emotion it must be part of an entire integrated situation.*

William James has given many examples of emotional experience drawn from a variety of sources. One which not only describes in small compass the experience of an intense emotion but also gives us some insight into the make-up of the emotion itself, is as follows:

> I remember the night, and almost the very spot on the hilltop, where my soul opened out, as it were, into the Infinite, and there was a rushing together of the two worlds, the inner and the outer. It was deep calling unto deep,—the deep that my own struggle had opened up within being answered by the unfathomable deep without, reaching beyond the stars. I stood alone with Him who had made me, and all the beauty of the world, and love, and sorrow, and even temptation. I did not seek Him, but felt the perfect unison of my spirit with His. The ordinary sense of things around me faded. For the moment nothing but an ineffable joy and exaltation remained. It is impossible fully to describe the experience. It was like the effect of some great orchestra when all the separate notes have melted into one swelling harmony that leaves the listener conscious of nothing save that his soul is being wafted upwards, and almost bursting with its own emotion. The perfect stillness of the night was thrilled by a more solemn silence. The darkness held a presence that was all the more felt because it was not seen (8, 66).[1]

Another example from the same source, but of a very unpleasant experience, is as follows:

> I had such a universal terror that I woke at night with a start, thinking that the Pantheon was tumbling on the Polytechnic school, or that the school was in flames, or that the Seine was pouring into the Catacombs, and that Paris was being swallowed up. And when these impressions were past, all day long without respite I suffered an incurable and intolerable desolation, verging on despair. I thought myself, in fact, rejected by God, lost, damned! I felt something like the suffering of hell. Before that I had never even thought of hell. My mind had never turned in that direction. Neither discourses nor reflections had impressed me in that way. I took no account of hell. Now, and all at once, I suffered in a measure what is suffered there (8, 146).[1]

Another description is of an emotional outburst of a monkey:

> Sometimes she gazed at the food and at the same time

[1] Reprinted by permission of Longmans, Green and Co.

> struck the floor with her hand or with a stick. Often she went around touching various objects and at the same time looked up at the food. In case repeated attempts to obtain the food failed she became more and more affectively disturbed. She went around violently striking the radiator, the floor, or a wooden box and excitedly uttering a number of sounds. While doing this she again and again looked towards the food. Finally she got into a state of what might be called a "generalized" affective disturbance, a state in which all reference to the goal was lost. She began throwing things around; she pushed and kicked various objects but she no longer threw them in direction of the goal; she did not even look at the goal. It seemed there was a diffuse discharge of energy instead of one specifically directed towards the goal (10, 298).[2]

Let us consider certain of the elements with which these examples provide us. In each instance the experience seems to overwhelm and bewilder the individual. He finds such extreme difficulty in adequately describing the experience when it is over that he usually resorts to analogy or to terse, exclamatory remarks. The name of the predominating tone may be constant or it may alter diametrically in a very rapid fashion. The experience pervades the entire personality. The individual feels this experience to be clear, true and undoubted. The intensity of the experience is in part made up of the total number of elements entering the situation, and in part by the degree to which the individual elements enter. As an entirety the emotional experience is at once overwhelming and satisfying, unique and many-in-one, pleasant and unpleasant, integrating and destructive, and healthy and unhealthy. However, it is always a part of an entire situation and to the best of our knowledge never exists alone or apart in human experience.

EXPRESSIVE EMOTIONAL REACTIONS

Since the reactions which occur during emotion are fairly distinct and may be observed and recorded, this portion of

[2] Reprinted by permission of the University of Chicago Press.

the field has received the most experimental investigation. Studies have been made of overt behavior, facial expression, various organic changes and the emotional reactions occurring during neuropathological states.

Experimental Studies. Recently there have been studies of the interpretation of bodily posture as an expression of emotion, which show clearly that certain postures are rather uniformly judged by observers to be expressive of some emotion or emotional reaction. Expression is judged on the *total observable material*; the concealing of the face, arms or legs not only modifies the expression which the observer feels is inherent in the posture, but sometimes even subtracts all 'emotion' from the material.

There have also been studies of the emotional responses of infants who were stimulated by hunger, dropping, restraint or pain. Motion pictures of the behavior of the babies were shown to students. The students named from 12 to 25 different emotions as appropriate titles for the reactions. Nurses and medical students, familiar with infant behavior, when shown the same photographs, named only seven or eight emotions. When the films were shown completely so that the actual stimulation was seen preceding the resultant reaction, there was a great deal more agreement between the judgments than in the other conditions, that is to say, the name attached to the response was strongly influenced by a *knowledge of the stimulus.* In another experiment the stimulus situations were attached in the films to behavior which had been elicited by other stimuli. These falsified pictures gave rise to interpretations which indicate that the stimulus preceding the reaction is usually the deciding factor in the name given to the reaction. There is little evidence for the belief that anger, fear and love are the three innate patterns of emotional response in infants. Emotional responses vary from child to child and also from one age level to the next. The important point is that these experiments do *not* show any *innate pattern* of emotional response which uniformly appears in the developing human organism (15).

In one experiment the investigator took more than 1200 photographs of the facial expressions of 25 subjects who were exposed to a series of emotional situations. Analysis of these photographs showed that each individual tended to use certain individual expressions (certain muscles or muscle groups) in practically all of his expressive reactions. Marked individual variations in these expressions existed. The involvement of no expression (group of muscles) was found to be significantly associated with any single type of stimulation or with any single type of verbal report of emotional experience aroused by the experiment. The most common facial reaction was smiling, which occurred in one-third of the photographs. The verbal reports of the emotional experiences indicated that pain, surprise, anger, exasperation, crying, disgust, sex and revolting experiences gave decreasing amounts of expressive movement in the order named. The fact that each individual had two or three expressions which appeared in the great majority of his responses to the varying situations and experiences indicates that the successful interpretation of emotional reactions by an observer must rest in part upon his familiarity with the usual reactions of the subject and in part upon a fairly complete knowledge of the entire situation.

There have been a number of other experimental studies of the conditions and reliability of the interpretation of emotional expression. There is little agreement among persons who are asked to interpret the actual crying of infants who are crying because they are hungry or because they have been dropped or restrained. Nor is there much consistency among the judgments of children who are asked to interpret the emotions conveyed by the sounds of phonograph records of the alphabet recited in tones of fear, anger, happiness, etc. The reliability of such judgments increases with the intelligence of the child making the judgment. In music the vibrato is found almost always to add an emotional coloring to the sound.

Since pronounced emotional reactions are usually accompanied by changes in organic or *visceral* functions, many ex-

periments have been conducted in an effort to ascertain the relationship which exists between cardiac and circulatory reactions, respiratory responses, the metabolic rate, gastro-intestinal reactions, etc., and emotional experience. Many of these experiments have had as their source the historically famous James-Lange theory of emotion.[3] In brief this theory stated "that *the bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur* is *the emotion* . . . that we feel sorry because we cry, angry because we strike, afraid because we tremble, and not that we cry, strike, or tremble because we are sorry, angry, or fearful, as the case may be" (7, 375). This theory has in general been discredited by the experiments which were designed to confirm it.

There is no need for us to consider here the vast amount of experimental research which has been undertaken in the hope of demonstrating a significant relation between visceral change and emotion. Only the most meager positive results have issued from it. It is true that there is a disturbance of circulation in emotion, but the nature of the circulatory change is not predictable nor does it differ significantly in different emotions. Emotion increases blood pressure, but the attempts to use this change as an index of conscious deception have proved unreliable. The correlation of respiratory changes with emotional states is equally unsatisfactory. The rate of metabolism is increased in emotion, but it affords no index by which emotions can be differentiated. Similarly the tone of the gastro-intestinal tract is changed in emotion without its being specific for a particular emotion. In strong emotion gastro-intestinal contractions may be inhibited. In brief the visceral changes in emotion lack the specificity which would render them useful in an understanding of the complexities of emotional phenomena.

All studies show that these organic responses are very closely interrelated. Excitement, startle or a mild increase in general

[3] The theory has been named after the American psychologist, William James, and the Danish physiologist, C. G. Lange.

activity usually causes an increase in visceral activity. Painful, strong, pronounced stimulation of any variety tends to interfere with the complicated integration of organic functions. Organic responses are generalized, lacking individuality of pattern with respect to the emotion which is being experienced. Patterns of organic response are peculiar to certain individuals. Certain people have one or another pattern of response which is usually associated with one or another emotional experience. Since these are associated, the tendency is for similar organic reactions in that individual to be interpreted as having a common experiential element. However, there is no evidence to show any similarity extending from one individual to another with respect to these patterns.

Smiling, laughing and crying. The most clear and well-defined cases of emotional expression are those of smiling, laughing and crying. We uniformly accept the occurrence of this sort of behavior as indicative of emotional experience. It is true that this behavior may occur without attendant emotional experience, but in the ordinary conduct of everyday life, smiling, laughing and crying are by common consent regarded as true emotional expressions.

The development of these patterns of response in the infant has been a matter of interest in child psychology. *Smiling* is exhibited at a very early age by most children. In the very young infant it is almost invariably brought about by specific stimulation, and the response is usually evoked by, or in the presence of, other people. By the end of the first year smiling has become a learned response to an extent which indicates that the smile is now a communicative, adaptive, *social reaction* rather than a purely expressive response. Laughing appears much later in the child's life, usually after the twentieth week. It is a stereotyped form of behavior during the first year of life. Individual differences between children occur mainly in the frequency of smiling or laughing and do not depend upon the form of this behavior. Smiling or laughing are not always different degrees of the same behavior pattern.

Laughing is an expressive emotional pattern even at its first occurrence; smiling is either social or emotional depending upon the circumstances (17).

The behavior of infants during *crying*, in a series of standard situations, has been observed experimentally. The frequency of crying induced by strangers increases up to about ten months of age. Crying, because of fear and strange situations, can be distinguished from other types of response. Smiling, laughing or crying in the adult is so bound up with the social reactions of the individual that it is impossible to be certain in a majority of cases whether the response is truly emotional, only partly so or devoid of emotion.

Adults have been observed in situations (funerals) which produce tears. The conclusion of this investigation is that tears are usually indicative of a *mixed* emotional state. Sorrow, dejection, joy and elation, when occurring alone, are but little effective in producing tears. Adult crying occurs when a depressing or otherwise unpleasant situation gains a redeeming feature, or when a period of tension with an unpleasant stimulation is followed by pleasant or alleviating stimulation.

Laughter involves a variety of psychological problems. We have the laughter of joy, laughing at the comic, laughing as a form of social phenomenon, laughing as a release of tension and laughing in pathological, organic or mental conditions. All of these involve different psychological elements. The joyful laugh, which is a bubbling over of good humor, is commonly shown by children or by adults when in a state of well-being. The comic laugh is directed at some joke or ludicrous situation. The entire problem of wit and humor is one which has held the interest of philosophers and psychologists for centuries. No one major line of explanation covers all laughing at the comic. The problem of laughter as a social phenomenon is one which involves all of social psychology. We laugh more easily in a group than when alone. Both laughter and facial expression are varieties of gesture language. The act of laughing is at times a communication of good will

and a spirit of fun, at others of pure joy and at still other times of embarrassment.

The laughter which is associated with a relief of tension has been explained on an evolutionary basis. Since the facial musculature is not primarily necessary to the active energetic preservation of life, the excess energy set up by emotional stimulation is drained off by the activity of the facial and respiratory musculature in a way which does not interfere with any other activity of the body which may be going on at the time.

In several forms of organic nervous disease the phenomena of uncontrolled and uncontrollable laughing and crying appear. Individuals will burst into gales of laughter or sob bitterly at the slightest stimulation. They may or may not experience amusement or grief while exhibiting this behavior. Some of the more intelligent patients will complain of their embarrassment at their behavior, since they recognize that it is inappropriate to the situation. That this behavior is associated with lesions in the thalamus has been claimed by the neurologist, Sir Henry Head. However, it has been shown that the behavior may accompany lesions of almost any part of the nervous system which may or may not involve the thalamus.

Anger and fear. The behavior of either animals or men, which by common consent is called *anger* and rage, is well marked and agreed upon by observers as a definitely interpretable form of expressive reaction. Actually the expressive reactions of anger and fear are but little different so far as the bodily responses are concerned. The real differences between these two presumably opposing emotional states rest in the total situation which arouses the states. Reactions which occur during offense, attack and positive movements are usually called anger; those which occur during defense and retreat are usually called fear. When the cortex of the cat or dog is removed without injury to the rest of the brain, the animal exhibits a series of reactions which has been called 'pseudo-rage.' The

teeth are bared, the animal claws indiscriminately, its hair is erect and it snarls or spits; but there is some doubt as to whether this is really *rage* or whether it might not better be called general excitement. In the child, fear and anger are sometimes studied from the standpoint of the *preparedness of the organism.* This preparedness or excitement is originally undifferentiated, but it becomes more distinctive and varied as the individual grows older, finally growing into socialized emotion, partly through maturation and partly through learning (9).

The stimulus condition for anger in young children is a situation which, instead of being a sudden call for action, is often a more or less sudden stoppage or *interference with action.* Interference with activity, especially activity motivated by the common urges or drives, is an essential criterion of the anger-producing situation. The anger responses in the child are outbursts of impulsive activity, such as kicking, stamping or slashing of the arms, and often prolonged holding of the breath. With increasing age the anger becomes more overtly directed toward a given end. The proportion of outbursts in which the behavior consists of displays of undirected energy decreases, while the frequency of retaliative behavior increases. The percentage of observable after-reactions, such as resentfulness and sulkiness, increases steadily with advancing age.

The stimuli for *fear* may be regarded as arousing responses to certain changes in the total situation, changes requiring a sudden new adjustment which the individual is unprepared to make. Among the changes which are fear-producing for the very young infant are those which substitute loud sounds or loss of bodily support for previous quiet, comfortable bodily balance. As the child develops and as his perceptual life becomes wider, things occur which startle or frighten him, because he perceives that they are new and unusual. Fear arises when so much is known to the individual that he recognizes the potential danger of a situation without having a complete apprehension or control of that situation.

The overt reaction for fear changes, in the growth of the young child, by the substitution of the more adequate specific response for undifferentiated panic. As the child grows older there is a development of a greater variety of responses. After the screaming, crying and rigidity the child adopts running away or partial withdrawal from the fearful situation. Exclamations and laughter are often substituted for crying at a sudden noise or fall. Gradually a more or less indifferent adjustment supersedes the panic.

Both anger and fear responses are easily attached to a new or different stimulus. These quickly established conditioned emotional responses may be altered by appropriate training or may persist over long periods of time. Fear may be unconditioned by appropriate associating of a pleasant stimulus with the fear stimulus. However, much work of practical scientific significance remains to be done on the unconditioning of emotional responses.

Various investigators have shown that the control of anger and fear is closely associated with the development of intelligence. The emotional outbursts of defective children correlate negatively with their intellectual level as measured by intelligence tests.

The adequate stimuli for anger or fear responses in the adult are very diverse in their forms of expression. The anger responses manifested by college students have been found to be complicated and variegated. "The causes or stimuli which arouse anger range all the way from a thwarting of a desire to do nothing to the interference with a desire to do everything. The impulses felt during anger range from a desire to injure, and even kill the offender, to serious self-injury, and from fight to flight. The responses during anger may be dominantly verbal, physiological, psychological, or social in nature and may range from a pleasant reply to doing violence to the offender. The after-effects range from a very reduced self-feeling to feelings of exalted self-importance" (14, 306). Strictly speaking there seems to be no such thing as a specific anger or fear response. *There are as many different manifesta-*

tions as there are configurations of behavior patterns and social situations.

Love reactions. Under this heading we may consider such diverse manifestations of behavior and experience as the parent-child relationship, the protecting reactions, the appreciation of protection, friendship, attachment between the sexes, attachment to places, to things, to occurrences or to food. According to Freud the central urge or drive of life is the *libido*, and this libido is wholly sexual in nature, although the term sex is applied in its widest sense. If Freud is correct, we have a reasonable explanation for the tremendous resistance to experimental investigation which is found with respect to these reactions. If life itself is so sexual and so intimately bound up with libidinal reactions, experimental investigation of the subject from a psychological point of view can be compared only to a physiological study of the reactions of the heart by means of human vivisection. It is self-evident that these responses are surrounded by a very extensive and formidable defense of social custom, law and taboo. That these taboos and customs are constantly being broken by the individual and even altered by a whole population has been shown by various studies of the sex life of the normal individual. These investigations have shown that practically every individual experiments with these 'love responses' outside of the realm of the so-called socially approved reactions. They also show that, for the most part, these experiments of the individual have but little permanent effect upon the total personality.

Since this general subject is so strongly guarded by social custom, practically all the experimental work in the field has been done upon animals. The studies which have been made on the development of patterns of sex behavior and maternal protection in the rat show that here, at least, we have a type of very complicated behavior which appears spontaneously and without learning, when the animal becomes sexually mature and when the first litter is born. By means of measurement of the amount of punishment (electrical shock) which an ani-

mal will endure in order to reach a desired object, it is possible to rank the strengths of these types of reaction in the rat (see pp. 393f., 456-459). The comparison of these drives does not give a relative measure of emotion, but the method offers the suggestion of a possible way in which the question might be approached in human behavior.

Other emotional reaction patterns. The reactions of startle or surprise are brought about by sudden and unexpected changes in the situation which is affecting the organism. The responses are usually the sudden stoppage of voluntary activity, a marked tensing of the musculature, verbal exclamations and rapid alterations of visceral rhythms. All these responses are of biological significance. The sudden pause and immediate defense reactions are of importance for the survival of the individual.

Pain is a *sensory* phenomenon, but it does set up many emotional responses. The physical injury which arouses pain leads to a diversity of responses, all of which are directed either to avoiding, removing or enduring the stimulation. Biologically, the emotional element in pain is probably associated with the complex conscious experiences grouped under the term of *self-preservation.* The association of painful stimulation with almost any type of previously neutral stimulation will lead to a conditioned emotional reaction, which becomes attached to the previously emotionally neutral stimulus. For this reason the phenomena of pain, although sensory in quality, complicate practically all the psychological reactions of an individual.

The experience of hunger or thirst results from the absence of food or water. Persistence of these sensations leads to a series of responses which may be termed emotional, since they partake in nature of all the criteria which have been used to describe emotional reactions. These responses are diverse in nature, are accompanied by expressive pleasant or unpleasant reactions and by marked changes in the usual visceral rhythms. Since the responses are complicated by a prominent and defi-

nite sensory element, they have not usually been regarded as of immediate emotional importance.

Other so-called emotional reaction patterns, such as disgust, contempt, pity, sympathy, jealousy, depression, have been listed by various investigators as emotional reactions or as emotions. Little or no evidence exists today which would indicate that there is anything unique psychologically in any of these experiences or responses. In every case it seems that the name is one which is assigned to some particular configuration or type of situation. However, the elements which can enter into the configurations are extremely varied, particularly with respect to their interrelationship.

NEURO-HUMORAL FACTORS IN EMOTIONAL REACTIONS

The attempt to establish an identity of occurrence of either emotion or emotions with some physiological reaction system or pattern has been a favorite line of research in physiological psychology for many years. The research which concerned itself with circulatory, respiratory, metabolic and gastro-intestinal changes has already been discussed.

In addition to the central and peripheral nervous systems which have been considered in previous chapters there is another integrative system which controls the visceral or vegetative functions of the organism. This is known as the *neuro-humoral* system. It is made up of the *autonomic* nervous system and the glands of internal secretion. Through the action of this biologically more primitive system those functions in the body which have to do with direct physical control of internal bodily integration are maintained.

The autonomic nervous system is essentially a nerve net of interconnections. Anatomically, it is divided into the *sympathetic* and *parasympathetic* divisions. The parasympathetic division is composed of the *cranial* and *sacral* divisions. A schematic diagram of the interrelations of these divisions and their relationship to the various bodily organs is shown in Fig. 125.

Generally speaking, the activity of the sympathetic division

is antagonistic to that of the parasympathetic. For example, the heart rate is inhibited by nervous excitation reaching it over the parasympathetic whereas it is accelerated by excitation from the sympathetic. Those portions of functional control

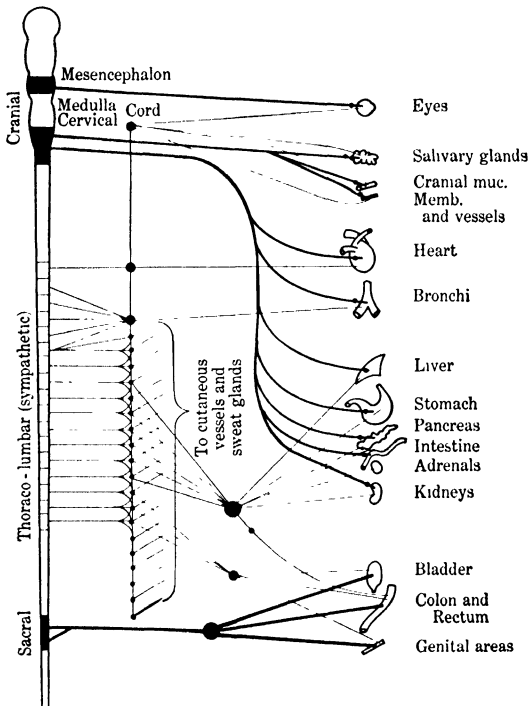


FIG. 125. SCHEMATIC DIAGRAM OF THE AUTONOMIC NERVOUS SYSTEM.

are not clear cut or exact, since it is complicated by the condition of the particular organs at the time of excitation, strength of the impulses, etc. Furthermore, the autonomic nervous system for the most part serves smooth muscle and glands so that its action is diffuse and relatively slow in effect in comparison with that of the central nervous system.

Activities which are initiated or inhibited by the action of the autonomic nervous system, *e.g.*, certain aspects of bodily

posture, gastro-intestinal tonus, heart rate, etc., are not, generally speaking, represented in consciousness. These actions and reflexes are involuntary. To the best of our knowledge they are innate and unlearned and probably modified only indirectly by any learning process. Their highest direct representation in the nervous system is in the *diencephalon* or interbrain.

Stimuli which bring about emotional reactions are particularly effective in their action on the autonomic nervous system and glands of internal secretion. In fact most of our remarks concerning the action of the autonomic nervous system are directly descriptive of much emotional reaction. The balance or lack of balance of visceral activity, the diffuseness of this activity, the involuntary nature of the responses, the lack of direct cortical representation, etc., all mark much of our expression of emotion.

We are not yet in a position to say that emotional reactions are uniformly identifiable with the action of the neuro-humoral mechanisms, since we are still profoundly ignorant of many phases of the true nature of both emotional reaction and of the neuro-humoral control. But as far as our knowledge goes at present there does seem to be good reason to believe that much of 'emotion,' as it is conceived of in the biological sciences, is mediated particularly by the action of the autonomic nervous system and glands of internal secretion.

There are in the body a number of secretory organs or tissues which throw their product directly into the blood stream. Such organs are called *endocrine* glands or glands of internal secretion. The *adrenal* glands are directly involved in the physiological expression of emotion. Two forms of secretion are known to be produced by these glands: that of the *medulla* of the gland which is called *adrenin*, and that of the *cortex* of the gland which is called *cortin*. Cortin has, among other things, the function of maintaining the oxygen-carrying power of the blood. It is absolutely necessary for the life of the organism. Emotional stimulation leads to the secretion of adrenin. An excess of adrenin in the blood stream brings about what

Cannon has termed emergency reactions of biological survival value.

Cannon and his co-workers have demonstrated that adrenin is secreted directly into the blood stream during emotional stimulation. An increase in the amount of adrenin in the blood has the following effects upon physiological activity: (1) it increases the tremor in voluntary muscle; (2) it causes relaxation of smooth muscle; (3) it counteracts fatigue in voluntary muscle by affecting the myo-neural junction; (4) it alters blood distribution; (5) it alters blood pressure; (6) it hastens the rate of clotting of the blood; (7) it relaxes the bronchioles; (8) it causes the liver to release glycogen into the blood stream; and (9) it causes the spleen to secrete or release red blood corpuscles into the blood stream. All these physiological changes which are brought about by adrenin may be considered emergency reactions which enable an organism to meet a situation which will call for the quick and probably prolonged discharge of energy. These reactions do not offer a pattern of response which enables one to differentiate emotions, but they do furnish a background for emotion in general (2).

Several investigators have studied the effect of the direct injection of adrenin into normal and abnormal adults. They found that many individuals reported merely the physiological disturbance, as perceived without any experiential emotional accompaniment. Other persons experienced what has been called a 'cold emotion.' That is to say, they had an experience which they felt was somehow emotional in nature, although no situation which would justify the experience was present. One subject termed the experience "a feeling like an accident waiting for some place to happen." A few individuals reported, subsequent to the injection of the adrenin, true, satisfactory emotional experiences. Sometimes this experience was characterized as anxiety, apprehension, fear or anger, but the subject could detect no difference between his experience and any other emotional experience which he had had. There was no difference between the reactions of normal and psychopathic persons in this respect.

The *thyroid* gland lies in front of the upper part of the trachea. Its internal secretion is known as *thyroxin*. There is no experimental work to show that thyroxin is secreted directly in response to emotional situations. However, it is probable that such secretion does take place. We know that in goiter (a diseased condition of the thyroid gland) a marked emotional instability of the patient occurs. Such patients overreact to practically any situation involving emergency or pain, and are in a state of heightened tension and irritability practically all the time. In other diseased conditions in which the thyroid gland is not functioning, we find a marked decrease or absence of emotional expression or responsiveness. The patient is stolid, stuporous or sluggish. It would seem that the thyroid secretion sets up a physiological background for emotional response. Whether or not this background functions in other than abnormal or diseased conditions, it is impossible to say at present.

The internal secretions of the sex glands are known to be directly connected with sexual desire and sexual experience. However, actual experimental work on this subject has never been carried out with human beings. It seems probable that the secretions of other endocrine glands, such as the pituitary, are involved in the emergency reactions in some fashion. Direct evidence is not at present available on this point.

On the basis of the studies of the effect of adrenin, the action of the sympathetic nervous system and the appearance of pseudo-rage responses during certain brain operations, Cannon has proposed an "emergency theory of emotion." Briefly stated, this theory holds that essentially emotion is a preparatory reaction of the organism, which is of biological survival value in times of danger. All of these physiological responses are associated with the biological preservation of the organism and with the defense of the organism against attack. The unique psychological experience or experiences in emotion he attributes to thalamic patterns of response or thalamic contributions to consciousness. This theory certainly has much to commend it and accounts for many of the physiological re-

sponses, although the attributing of the unique element in emotion to thalamic function is not as convincing as it might be. (See pp. 394f. on the possible function of the thalamus for pleasantness and unpleasantness.)

MEASUREMENTS AND TESTS OF EMOTION

The general object of science is to secure measurements or adequate descriptive criteria of natural phenomena. To this end much research has been conducted in an attempt to obtain either adequate criteria or measurements of the expression or experience of emotion. Much of this work has been centered upon the organic reactions of the circulatory, respiratory or gastro-intestinal systems or metabolic changes. As we have seen above, there is no evidence that these changes are directly associated with either the degree or variety of emotion.

In 1888 the French scientists, Vigouroux and Féré, called attention to the fact that, when electrodes are placed on the skin and attached to the proper electrical measuring instruments, variations in the electrical properties of the skin appear from time to time. They also showed that during emotional experience or excitement there is an increase in these electrical variations. These electrical changes have been named the *galvanic skin response*, and the phenomena have been extensively studied by various physiological and psychological investigators. The response was first called prominently to psychological attention by the work of Jung and his pupils, who came to the general conclusion that the galvanic skin response is associated with repressed emotional complexes. They did not particularly consider the question whether this electrical response is associated with physiological and psychological occurrences other than emotion, and claims have been accepted more or less uncritically by many psychologists. More recent critical investigations show that these electrical responses usually occur during emotional experience, but that they also occur together with practically every other variety of psychological experience. The degree of electrical change is not

associated with the amount or degree of emotion experienced by the individual. In general, electrical responses of the skin occur at the same time that emotional experience occurs, but they also occur when conscious reactions of a non-emotional sort are taking place (11). (See pp. 390f.)

An entirely different method of measurement and test is the one which makes use of the questionnaire or rating scale, as, for example, the Pressey X-O test. This test consists of three lists of words. In the first list the individual is told to cross out everything he thinks is wrong; in the second list, everything about which he has ever worried; and in the third, everything he likes or is interested in. He is also told to encircle the crossed-out word in each line which he considers to be the worst, the most worrisome or the most interesting, as the case may be. The total number of words crossed out is called a *score of emotionality*, since, theoretically, the more things a person dislikes, worries about or likes, the more general emotionality he has. The encircled words are compared with a standard list which gives the most frequently encircled word for each line, and the number of encircled words which deviate from this standard list is the *score of idiosyncrasy*. Various investigators have reported that students who obtain high scores of emotionality and idiosyncrasy tend to have more than the usual number of emotional conflicts in school. In general, however, investigations of emotion or emotionality by the technique of the questionnaire or the rating scale have not given results which can be validated or which have been found to be of particular practical use or significance.

PATHOLOGICAL REACTIONS

In the various types of nervous and mental diseases we find exaggerations of what seem to be real emotional experiences or expressions. In *involutional melancholia* or in the depressed phase of the *manic-depressive* insanity, the prevailing affective tone or experience is that of extremely unpleasant depression, unhappiness, anxiety and apprehension. Such an individual is experiencing, as far as anyone can tell, a truly conscious emo-

tional state which is persistent and which he himself finds to be in no way different from the equivalent experience in his life before the disease started. In certain phases of general paralysis of the insane or in the manic episodes of manic-depressive insanity there is a marked *euphoric* condition. When in this condition the patient is unreasonably happy, pervaded with a sense of well-being and buoyant hopefulness and is unable to appreciate the depressing events occurring in every-day life. Except in degree this condition seems to be no different from the normal experience of excessive happiness. In neither the depressed nor the euphoric condition is there any marked, constant physiological state which differentiates this emotional status from the physiological state of individuals who are not experiencing this condition.

In Wilson's disease there is exhibited an uncontrollable laughing and crying. This behavior may or may not, as we have noted, be accompanied by an appropriate emotional experience. In Dejerine's disease there is frequently an experience of an intense pleasantness or unpleasantness which accompanies stimulation applied to an otherwise paralyzed portion of the body. This pleasantness or unpleasantness is referred by the patient to the stimulation, but the experience itself is unreasonably intense to an extent which is practically unbearable.

Are these pathological reactions true emotional experiences or responses? There is every reason to believe that they are. The person who has the experience is unable, for the most part, to distinguish it from any other emotional experience which he has had in his past life. The laughing and crying in Wilson's disease are truly laughing and crying. The actual physiological and psychological basis for these pathological responses is at present unknown, but there seems to be no reason to doubt their psychological reality.

THE EXPERIENCE AND THE EXPRESSION OF EMOTION

There is no known constant relationship which runs through

all of the diverse phenomena which are commonly labeled 'emotion.' We are fairly certain today that many phenomena of human life which are usually considered emotional do not really belong in this category, either psychologically or physiologically. All the evidence goes to show that the experience and the expressions of emotion need not go together nor need they have any essential relation to one another. The expression and the experience may occur independently, although as a usual thing they occur together. All these experiences which are classified as emotional are easily conditioned so that they become detached from their original exciting stimulus and attached to new and different stimuli. In this process of conditioning, reconditioning and unconditioning we have a great deal of social facilitation and inhibition. Most emotional expression and a great deal of emotional experience depend upon the social conditions existing at that particular time of the reaction, and it is impossible to separate the emotion from the situation.

REFERENCES

1. Bard, P. Emotion: I. The neuro-humoral basis of emotional reactions. In *A handbook of general experimental psychology.* Worcester, Mass.: Clark University Press, 1934. Chapter 6.
2. Cannon, W. B. *Bodily changes in pain, hunger, fear and rage.* Second edition. New York: D. Appleton Co., 1929.
3. Crile, G. W. *The origin and nature of the emotions.* Philadelphia: W. B. Saunders Co., 1915.
4. Dickinson, R. L., and Beam, L. *A thousand marriages.* Baltimore: Williams and Wilkins, 1931.
5. Goodenough, F. L. *Anger in young children.* Minneapolis, Minn.: University of Minnesota Press, 1931.
6. Hamilton, G. V. *A research in marriage.* New York: A. and C. Boni, 1929.
7. James, W. *Psychology (briefer course).* New York: Henry Holt and Co., 1892. Chapter 24.
8. James, W. *The varieties of religious experience.* New York: Longmans, Green and Co., 1902.
9. Jones, M. C. Emotional development. In *A handbook of child psychology.* Second edition. Worcester, Mass.: Clark University Press, 1933. Chapter 6.

10. KLUVER, H. *Behavior mechanisms in monkeys.* Chicago: University of Chicago Press, 1933.
11. LANDIS, C. Psychology and the psychogalvanic reflex. *Psychol. Rev.*, 1930, 37, 381-398.
12. LANDIS, C. Emotion: II. The expressions of emotion. In *A handbook of general experimental psychology.* Worcester, Mass.: Clark University Press, 1934. Chapter 7.
13. LASHLEY, K. S., STONE, C. P., DARROW, C. W., LANDIS, C., and HEATH, L. L. *Studies in the dynamics of behavior.* Chicago: University of Chicago Press, 1932.
14. MELTZER, H. Students' adjustments in anger. *J. Soc. Psychol.*, 1933, 4, 285-309.
15. SHERMAN, M. The differentiation of emotional responses in infants. *J. Comp. Psychol.*, 1927, 7, 265-284, 335-351.
16. SYMONDS, P. M. Diagnosing personality and conduct. New York: D. Appleton-Century Co., 1931.
17. WASHBURN, R. W. A study of the smiling and laughing of infants in the first year of life. *Genet. Psychol. Monog.*, 1929, 6, 397-537.

CHAPTER 17

ACTION

Man is seldom, if ever, quiet in his waking moments, and he is much less tranquil than was formerly supposed even when he is asleep. In response to stimulation he is constantly making movements, though often they are hardly noticeable. There are the incipient movements of his vocal organs and other muscles while he is thinking, the more overt response of the ever-present eye-wink, the shifting of his limbs, the restless movements of his body, the frequent turning of his head and the more coordinated activities, such as walking, talking, piano-playing and tennis. We are dealing here with behavior directed primarily toward a manipulation and understanding of things of the external world. This behavior which is dependent, for the most part, on the striped muscles is chiefly the concern of the psychologist. There is, in addition, the action of the smooth muscles, of the endocrine glands and of the muscles connected with the function of nutrition and of reproduction, but these movements are mainly of interest to the physiologist.

The importance of the behavior which the psychologist studies need hardly be emphasized. Behavior was probably present in the evolution of life before the development of a nervous system, if we may judge from the lower forms of life, such as the sponge, whose muscles are stimulated by direct contact only. It is the means by which the organism becomes adapted to the ever-changing external situation in order to survive. We have already seen (Chapter 13) how one learns to coordinate one's movements to meet these situations. Our present task is to explain the characteristics of these responses and how they are initiated.

This chapter was written by H. S. Langfeld of Princeton University.

DEVELOPMENT OF ACTION

Mass movement versus specific response. In order to understand the nature of behavior in adult life and to be able to form an intelligent opinion of the relative importance of *inheritance* and *experience*, it is necessary to know something of the development of prenatal behavior and of the behavior directly after birth.

The most general question for us to consider is this. Does behavior start with units of discrete and unrelated reflexes, or is there a general pattern of response out of which these reflexes are developed? Coghill's experiments upon the salamander clearly indicate an answer for this form of life. He found that the first movement of the salamander is a bending of the head either to the right or to the left, a movement which is produced by the contraction of the muscles behind the head. With the bending of the head there is the start of a muscular contraction which extends rapidly towards the tail and involves finally the whole animal, causing it to assume the shape of a coil or a C. A movement of the head in the direction opposite to its first movement follows before the movement caused by the initial bending of the head has been completed. The animal thus assumes an S-shape, and this S-shaped movement, when speeded up, becomes the typical swimming movement. The successive stages of the development of this movement are shown in the three pictures of Fig. 126. This movement of the whole body may be termed integrated gross movement.

The limbs of the salamander appear after the swimming reaction has developed. At first both sets of its limbs move with its trunk and only gradually do they acquire independent movement. When a limb begins to move independently, it first moves as a whole; then the movement at the elbow appears, and finally movements of the wrist and digits gain independence. The conclusion which Coghill drew from these observations is that there is at first a total pattern or integrated

unit of behavior and that it is from this pattern through individuation that finer responses emerge as relatively independent units (5).

Experiments upon the fetuses of guinea-pigs, rats, cats and other animals have produced results which seem to sub-

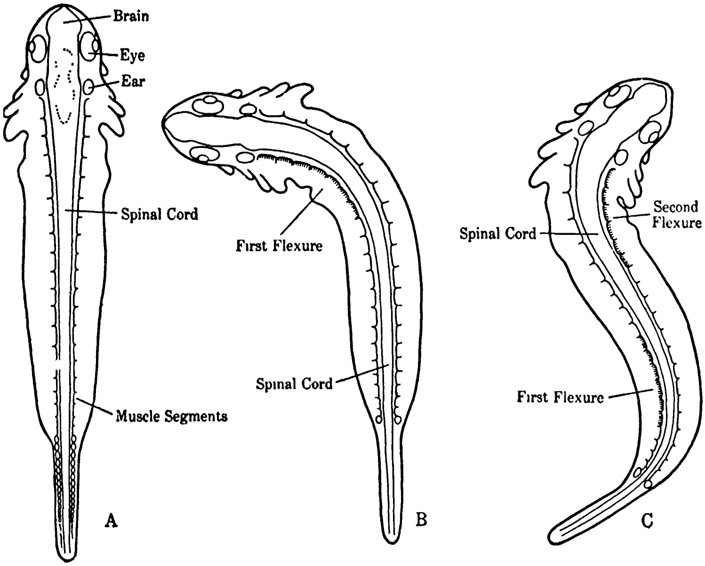


FIG. 126. SWIMMING DEVELOPMENT OF THE SALAMANDER.

A, stage without evidence of muscle contraction. *B*, beginning of swimming. C-shaped movement. Flexure is produced by contraction of anterior muscle segments, indicated by cross-hatching. *C*, S-shaped swimming movement. The first flexure has progressed toward the tail; the second flexure has started in the region of the head. Muscle contractions are indicated by cross-hatching. After G. E. Coghill (5). Reprinted by permission of the Macmillan Company.

stantiate this general principle. Although the experiments upon the human fetus do not give as clear-cut data, there seems to be the general tendency for stimuli at first to spread, causing gross movements, and for various individual responses to appear gradually during the developmental process and to become more or less specific according to their nature and circumstances. These gross movements are frequently termed 'random' from the fact that, as yet, they do not seem to have

any particular goal—as, for example, the uncoordinated movements of the arms and legs that occur when various parts of the skin are stimulated. It is interesting to note, in this connection, that even the adult, under painful stimulation, frequently reverts to this primitive form of response. He 'writhes in agony.' It would be incorrect to suppose, however, that, even in the first stages of development, stimuli always produce a general movement of all muscle groups, as if the human nervous system were in the beginning similar in action to the nerve net of the sea-anemone. There is merely a tendency at first for the stimulus to spread and for it later to be restricted to more definite forms of response.

Heredity and environment. The earliest movements of the human embryo take place in the second month of fetal life and consist of slow uncoordinated movements of the trunk and limbs. Reflexes are reported to occur in the third month, but it is difficult to decide whether they are true reflexes. However, at the end of the third or the beginning of the fourth month, a response of the toes has been produced by stimulation of the sole of the foot. During the fourth month many responses become more specific and more intense. In the sixth month certain tendon reflexes are known to occur, and in the seventh month the knee jerk. From this time the increase of specific responses is marked. There is also evidence that the proprioceptors function early in fetal development, thus initiating impulses for muscular contraction. In short, the human fetus has already an elaborately coordinated and organized response mechanism at the time of birth, one which forms the basis for the fundamental behavior of its later years (3).

The origin of these patterns of response is not exclusively in the chromosomes of the fertilized ovum. It is more likely that their development is in consequence of a *constant interplay between the organism and its surroundings.* Heredity plays a great part in determining structure; and maturation, or growth, which is due to the interaction of the various parts

of the organism, is also an indispensable factor. It is principally through this concept of maturation that one can explain the appearance of new activities at various age levels. There is, in addition, the very important effect of learning in adaptation to the environment. The organism has accumulated a considerable fund of experience before birth. Learning commences at the time of the first movement in the uterus, and from then until death the process of habit formation through the establishment of more or less fixed patterns of response is continuous.

The relative importance of these various factors of development is very difficult to determine, since it is through their constant interaction that the development takes place. No doubt exists, however, that, at least in lower forms of animal life, there are situations in which heredity and maturation play a dominant role and where the environmental influences are at a minimum. Experiments have been conducted where the animal was rendered incapable of overt action, and could therefore not gain any experience. Embryos of the frog and salamander were placed in an anesthetic long before they reached the stage at which movement can be observed. As a check the experimenter placed other embryos in fresh water. At the time when these latter were swimming freely about, the anesthetized organisms were still inactive. When, however, the anesthetized embryos were placed in fresh water, they soon showed signs of movement upon stimulation. It is difficult to observe these first movements, but the experimenter described them as "twitchings or slight movements of the head or tail, such as one usually denotes as random" (4). In a few minutes they were swimming so perfectly that they could not be distinguished from individuals of the control group.

These results show that, at least in such lower forms of life, the neuro-muscular mechanism can develop normally without any apparent external stimulation and without response, and when it has sufficiently matured will produce relatively integrated movements almost the first time it is appropriately

stimulated. We apparently have here an example of the minimum amount of environmental influence compatible with normal development. Studies of various other forms of life, however, seem clearly to indicate, not a decrease in the effect of heredity, but an increasing influence from the environment in the progression from the lower to the higher forms of life until one comes to man, where environmental influence seems to be at a maximum.

Behavioral development at birth. Animals differ greatly in the degree of their behavioral development at birth, just as they differ in physical maturity. The rat is born without hair and is scarcely able to crawl. It lies with its litter mates, a helpless limp form, almost entirely dependent upon its mother. The kitten, though not quite so helpless or so immature physically, does not walk perfectly until the fourth week. The guinea-pig, on the other hand, is born with an adequate coat of fur and walks and runs almost at once. The colt and the calf also have well-developed action patterns at birth and are able to stand and walk almost immediately. A class of animals that shows an unusually early development of behavior is the marsupial, to which the opossum and the kangaroo belong. These animals are born in a condition which may be considered that of a very immature fetus and are cared for by their mother in a pouch where the nipples are situated. They travel from the vulva to the pouch and attach themselves to the teats entirely unaided by the mother, in spite of the early stage of embryonic development. The opossum, for example, when it makes this journey, is only ten days removed from an unfertilized ovum.

The human infant at birth does not rank very high so far as its actions are concerned, and it takes a long time to develop those habits which will make it independent of its mother. Yet, as we have seen, it already has a great many simple modes of response which have developed during the prenatal state, and which are the basis of the more complex behavior to follow. It is able to make elementary movements of its arms and legs and trunk. It attempts to lift its head and

rear quarters and is very soon successful. It stretches by bending the head and extending the hands above the head. It makes 'startled' responses in which the arms are moved apart, fingers spread, the legs extended and the head thrown backward. Smiling may occur on the first day of life if the infant is tickled under the chin. Crying occurs at birth and is often accompanied by activity of the arms and legs. Various stimuli, such as hunger, a bright light, a loud noise, falling, cold, heat or pain may cause crying. A pinprick on the hand will cause a withdrawal of the arm, and a slight push on the sole of the foot is sufficient to cause an extensor thrust, which often has sufficient strength to support the infant's weight.

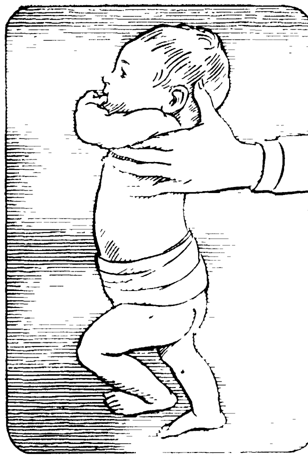


FIG. 127. EARLY STEPPING MOVEMENTS.

Infants a few hours old when supported under the arms frequently make prancing steps which are probably the early stages of walking. Sketch of a photograph from McGraw, M.B., from reflex to muscular control in the assumption of an erect posture and ambulation in the human infant, *Child Development*, 1932, 3, p. 295.

There are also more highly coordinated movements at birth. When the infant is placed prone, the arms and legs are drawn under the body and frequently each pair makes reciprocal movements as in creeping. When it is held upright with its feet touching the floor, it makes stepping movements. The grasping reflex is also present at birth. If a stick is placed on the palm of an infant's hand, its fingers will curl about the stick and it will hold on with considerable strength. Indeed, shortly after birth one is able to raise an infant from the ground by its hold on a stick.

QUALITATIVE DEVELOPMENT OF ACTION

As we have seen, these responses which appear at birth are not due solely to inherited factors. One explanation of the de-

velopment of such activities as the grasping reflex is given in the theory of E. B. Holt. This theory is based on the principle of the *reflex circle*, a principle which states that, when a stimulus causes a sensory impulse at the same time or immediately succeeding a motor impulse, there is a very strong tendency for the sensory impulse to find an outlet along the motor path which is still in a state of excitation. For example, if an infant utters the sound "ah," this sound stimulus affects its ear and an impulse travels along the auditory nerve. Since the muscles of the vocal organs are the ones that have just moved, it is along these motor paths that the impulse derived from the auditory nerve travels, and again the infant says "ah." It is clear that until there is a break in this circle, the infant will continue to utter the sound—a phenomenon that is frequently observed.

In order to see further how this principle works we may take Holt's explanation of the grasping reflex. He writes:

> Consider a random impulse reaching the flexor muscle of a finger. In the foetal position the fingers are often closed over the palm of the hand, and the least random flexion of a finger will cause it to press on the palm. Then (what is not random) afferent impulses ("tactile") from the two surfaces in contact (palm and finger) will be sent back to the central nervous system, where by the principle already cited they will find an outlet in the motor paths that were just now excited, that is, those of the flexor muscle of the finger in question. When this has happened a few times (as it is bound to happen) the reflex-circle will be established; and then a pressure stimulus on either palm or finger will cause the finger to flex and so to close down on the object that caused the pressure. Such is the origin of the "grasping reflex," which is so useful through all the later life (7, 38f.).

Specialization and Generalization of Response. In the development of behavior there is, as we have seen, a marked tendency for responses to become increasingly *specific*; the organism, in other words, tends to become more delicately and adequately adjusted to varying situations. This adjustment

can take place in several ways. The stimulus may become more specific, the area of stimulation may become more restricted or the response may become more specific. All three directions of specificity are generally involved in the course of an adjustment.

An experiment on the conditioned response illustrates the development of specificity of stimulus and of locality of stimulation. A tone of a definite pitch is sounded just before a dog is fed. This stimulation by tone and food is repeated a number of times. The tone is finally given without the food and there is a discharge of the salivary gland such as occurred when the food was presented. Now the point for our present purpose is that, in the early stages of training, there is likely to be a salivary response, not only to the tone which was used, but also to tones differing considerably in pitch from the original tone. Eventually, however, the gland responds only to the tone used in the conditioning process or to one differing only slightly from this tone. Here we have plainly obtained a specificity of stimulus.

The development of specific localization of stimulation is shown in the experiment in which a dog is touched on a certain spot on the side and immediately afterward given an electric shock on the leg sufficiently intense to cause a flexion of that member. After a number of trials, touching the side alone without the electric shock will produce a flexion of the leg. However, in the early stages of conditioning, one may obtain the leg movement by stimulating any point of the skin within a certain area around the point originally stimulated. This area becomes increasingly restricted with repetition, until finally the response of the limb is elicited only when the originally conditioned spot is stimulated.

The progress toward greater specificity in the response of the organism is evidently present in all learning. The rat in learning to run a maze gradually eliminates useless movements into blind alleys until it can reach the goal without errors. The typist at first makes many false finger movements. In

learning to play golf, one is apt in the early stages of learning to move more than the specific muscles necessary for the swing, especially when an unusual effort is made; and this tendency reappears, even after long practice, when one becomes fatigued—as is evidenced by the moving of the head and swaying of the body.

Although these fixed modes of response to definite environmental situations are necessary if the organism is to become properly adjusted, obviously this process of specification alone would produce mere automatism. Man soon learns also to vary his response to a stimulus according to the total situation.

Generalization parallels specialization in all but the lowest forms of intelligence. We have only to consider the genesis of the language response to realize how true this statement is. The normal child not only learns to react specifically to such words as "da-da" for its father, but eventually to act differently, according to the circumstances, toward general terms, as for example, to the word "man." At first it probably reacted with the word "da-da" to all the men it saw, but this could hardly be called a general response; rather it is a specific reaction to a series of varying stimuli whose differences it had not yet learned to recognize—a form of response which the mentally defective child may never learn to alter. When the child uses a term like "man" to a kind of human male it had never seen before, then one is justified in calling the response a generalization. (See pp. 489-493 for account of generalization in thought.) In fact, it is this generalized response which is fundamental to intelligence, according to the definition of intelligence as ability to meet new situations on the basis of past experience.

Reflex action. If liquid gets into the throat of an infant, its swallowing muscles respond. If there is too much liquid, it chokes. The infant starts breathing at birth. Its eyelids close at a loud noise or a movement toward the eye. Its glands commence functioning during the fetal period. There are many other movements of this nature in the human repertoire. They

are called *reflexes* and may be defined as involuntary and prompt responses of the striped or of the smooth muscles.

The simplest form of reflex would require a receptor, sensory neuron, motor neuron and effector. Such a simple neural arc, however, is not found isolated functionally from all other parts of the nervous system in a mature human organism. Take, for instance, the following example of a spinal reflex. If one pinches the paw of an animal whose cord has been cut just below the medulla, one can still obtain a withdrawal or flexion of the paw accompanied by a forward thrust of the other paw. A relatively simple neural arc is involved in the flexion of the paw, but even here more than one motor neuron is necessary to bend the leg; and besides there must be a connection in the cord between the sensory neuron and the motor neuron going to the opposite leg to produce the thrust in that leg. There are also connections between these arcs and many more remote reflexes, which may exert either an inhibitory or facilitating effect upon the first reflex, if they are stimulated at the same time. Furthermore, if the central nervous system is intact, the legs may be moved voluntarily, a fact which means there are connections between the spinal reflex arc and the cerebrum. This brief sketch of the physiology of the *reflex arc* is given to emphasize once more the fact that a simple form of response is involved in a very complicated neural and muscular pattern.

For the most part, reflexes are unlearned in the broad sense of the term, although there is a question whether some external stimulation is not always necessary for their development. Certainly, in the case of such reflexes as the grasping reflex (see pp. 427f.), there appears to be a developmental history involving more than maturation. After the reflexes are once firmly established, they are stable and predictable, and many of them are invariably found in organisms of the same species.

That reflexes as such are *involuntary* does not imply that they are always *unconscious*. Many of our reflex actions occur without our being aware of them, as for example the reflexes connected with the functioning of the internal organs. Many

of them are usually unconscious, such as breathing and eye-winking, but they may become fully conscious if we attend to them. Of others we are perhaps always conscious, such as the knee jerk or the rapid withdrawal of the hand from an electric shock. Moreover, reflexes differ as to whether they can be controlled, as to whether, although ordinarily involuntary, they can be voluntarily initiated or inhibited. The reflexes which are always unconscious are for practical purposes beyond our power to regulate, but we can stop breathing or change its rate; we can wink voluntarily and can frequently prevent winking. We can indirectly even change heart rate by voluntarily recalling some scene highly tinged with emotion. Who cannot, for example, cause his heart to palpitate by voluntarily imagining himself charging over the top in battle?

It is even possible by the method of conditioning to obtain voluntary control of what is for most persons never anything but an involuntary reflex. In certain experiments of this nature, the pupil of the eye was trained to contract at command. In the first stage of the training, a bell was rung immediately before a light was shone in the eyes. After some trials the sound of the bell alone caused the pupil to contract. Then the subject was instructed to close and open the circuit for both bell and light by closing and opening his hand at the verbal command of the experimenter. The verbal command was thus connected through the hand movements and the bell to the pupillary reflex. The next step in the experiment was to eliminate the hand movements and the bell. This left only the vocal instruction of the experimenter as the conditioned stimulus to which the pupil now contracted. The last stage was the repetition by the subject himself of the verbal instructions, first out loud, then by whispering and finally subvocally. It was found that all these forms of stimuli became the conditions of the contraction of the pupil. The subject could, in short, now command his pupillary reflex, and this ability was still present after fifteen days without further conditioning in the meantime (8).

Instinct. Reflex action is frequently termed instinctive because it is, in the first instance, an unlearned response depending upon innate connections in the nervous system. The primary responses to the internal drives of the organism, such as hunger and sex, also fall under this category, and these drives in turn may be termed the motivation for such responses. More complicated responses are also called instinctive when they involve innate reflexes in their response patterns, and when organic needs or drives are their immediate causes. These complicated responses owe their development to experience as well as to innate connections, and it is the interest of the psychologist to endeavor to determine by observation and experimentation how much may be rightly classed as instinctive and innate and how much as acquired. It is for this purpose that observation and analysis are made of the various forms of both prenatal and infantile behavior. Much study has also been given to the adaptive behavior of animals.

An interesting habit, which is in part instinctive, is the pecking response of chicks. The chick moves its head somewhat in the manner of pecking before it is hatched. Owing to an internal stimulation the whole body moves violently in the shell and the head movements can be seen to take on the form essential to pecking. It is during one of these agitated movements that the shell cracks open and the chick emerges. The chick's action in breaking out of the shell is instinctive in the sense that it is caused by the internal development of the organism, but not in the more popular sense that the *idea* of getting out of the shell at the right time was inherited by the chick. After the chick is thus released it gradually learns to eat. At first it often misses the grain of corn that it strikes at. It may strike the corn but not seize it, or it may seize it, but not swallow it. Only after some days does it peck accurately and eat with the proficiency of the adult hen. If some of the chicks are fed artificially for several days and not allowed to peck during that time, they will nevertheless very soon learn to peck as accurately as the chicks who had started 'practicing' earlier. Thus we see, even in this relatively simple response of

pecking, that instinctive response, maturation and habit, all three, play a part.

Many animals build nests according to a pattern which varies little within the species. In some instances the offspring have had no opportunity to learn from their progenitors. Then at least there must be an innate tendency running through the activity. That such behavior, however, cannot possibly be an instinct in the sense in which an instinct is sometimes defined (*i.e.*, a series of chained reflexes whose connections are innate and determined) is evident from the fact that the animal must suit its behavior to the particular surroundings in which it finds itself and the kind of material immediately available for its purpose.

In other cases, so-called instincts, both in animals and man, are found to be learned behavior. Naturalists frequently have reported, for example, that the wild animals they have met were not 'instinctively' afraid of man until they had had unpleasant experiences with him. When a child touches a hot radiator it instinctively pulls its hand away—instinctively because the strong stimulus causes a withdrawal response in the muscles without previous training. When later, at sight of the radiator, the child avoids the painful contact, such avoidance can no longer be legitimately called instinctive even though it is immediate and unreflecting. It is a wise and prudent principle when explaining behavior to endeavor first to determine all the factors of experience that could possibly have been operative in the development of such behavior before concluding that the behavior is innate.

Differences between reflex, conditioned response and voluntary acts. In many instances of human behavior there is no difficulty in distinguishing a simple reflex from the more complicated conditioned response or from a voluntary response. For example, one knows whether he has winked voluntarily or whether the eyelid reflex has been caused by some sudden stimulus. In the conditioned patellar reflex, when the knee jerk is conditioned to a bell, one can distinguish the

movement of his leg at the sound of the bell from the movement which is caused by a rap on the knee, or one that is produced voluntarily. In these cases the person making the movement draws his inference primarily from the nature of the antecedents to the movement.

When it is necessary to decide merely on the basis of the overt behavior to what type of action a response belongs, a judgment is more difficult to make. In the case of the conditioned knee jerk, for example, the investigator cannot always tell whether the movement for the bell alone is involuntary or whether the subject is 'faking' results by voluntarily moving his leg when he hears the bell.

Numerous experiments have been directed toward obtaining some objective criterion, and it has been found that, in general, speed of reaction differentiates these three forms of response. The reflex is, on the average, the most rapid, the conditioned response is slower and the voluntary response is slowest. The experiments in which the pupillary light reflex was conditioned showed that the average latency of the conditioned dilation of the pupil was 1.56 sec. and of the conditioned contraction 2.29 sec., whereas the unconditioned response is generally from 0.2 to 0.5 sec. The duration of the conditioned dilation response was 8.24 sec. and of the conditioned contraction response 10.93 sec.; that of the unconditioned light reflex is usually from 1 to 4 sec. (8). There may be, however, overlapping. In the case of the eyelid response it was found that, through practice in opening the eyes as quickly as possible immediately after the eyes had closed, the speed of such voluntary opening increased above the speed of the reflex response. This result does not mean, however, that the voluntary response has developed into a reflex.

Other objective characteristics remain to differentiate the two forms of action. For example, if the total time of the wink is analyzed into the time of opening and the time of closing the eye, it is found that as the time of closing decreases, that of opening decreases. In voluntary response this relationship is changed (11). Furthermore, voluntary response is more

readily modified by instruction than is the reflex, and under certain conditions the change is in the opposite direction. For example, subjects have been told to relax as much as possible during both voluntary and reflex action. When the records of the lid movements are analyzed, it is found that the latency of response (the period between the presentation of the stimulus and the beginning of the movement) is slightly decreased for the reflex and increased for the voluntary response (10). These last results are readily understood. The football player has to be 'keyed up' to start immediately upon the snapping of the ball. If he relaxes for a moment he will be caught off his guard. On the other hand, the reflex seems to work best when one is caught off his guard. If one's attention is concentrated on the appearance of the stimulus for, let us say, the knee jerk, there is likely to be a slight tendency to inhibit the reflex.

Experiments have also been made to determine whether any differences between reflex and voluntary activity can be discovered in action currents. Thus it has been found that the pattern of the action current is more stereotyped in the reflex, a discovery which is in accord with the conception of a reflex as a *fixed* form of response as compared with the *variability* of voluntary response (12).

Voluntary and semi-voluntary acts. Although most of our movements in the early stages of life are of the reflex type, the voluntary, semi-voluntary and automatic acts soon assume the ascendancy. A voluntary act is a fully conscious one. It is an act in which one knows what he desires to do and how to make at least a beginning toward that end. When one starts to learn some difficult movement, such as a new kind of dive, he has an idea of the form of movement that he wishes to make, and especially is he conscious of the first movements involved in the spring from the board. If he is being taught, it is quite probable that he will repeat the instructions to himself subvocally as he stands on the board. During the dive he will be aware to a certain extent of the position of the

limbs, and after the completion of the dive he will have a memory of what he has done, accompanied by a certain consciousness of the amount of success attained. Or in learning to play the piano, one at first thinks of the finger movements connected with the note and the position of the key before making the necessary movement. We need not, in this place, describe the learning process, but we must note how fully conscious we are of what we do immediately preceding, during and after acts which we try to perform for the first time.

This type of voluntary action also appears when a person makes a mistake during the learning process, when, for example, he presses the wrong key in learning to use the typewriter. In such a case, even when he is already fairly proficient, he is apt to revert to the first stage of the learning process and to become conscious of the direction in which he should move the finger. The interruption in the smooth sequence of movement brings the movement again to consciousness. The reverse is also true. If one becomes suddenly fully conscious of what he is doing, when he is performing some well-coordinated response, there is apt to be an interference in the smoothness of the response. Let him be very anxious not to make a mistake in the letter he is typing and he is almost sure to do something wrong. If one thinks of voluntarily moving his legs when going rapidly upstairs, he is likely to trip. By changing the automatic response to a voluntary act, one throws it back to the initial stages in the development of the habit.

Obviously, no one is completely conscious of the acts performed in the round of daily duty. A pitcher, when he throws a ball, does not have to think of the movements he is going to make. The act is voluntary in the sense that he intends to pitch the ball, but, as he starts the swing of his arm, he is probably looking at the plate and his mind is occupied with little else than the corner of the plate he wishes to 'cut.' Also, during the act of throwing he is hardly aware of what his arm is doing, for his attention is still focused on the batter and the plate. It is evident that we cannot call such an action fully

voluntary in the sense that we should apply the term to the same pitcher's acts when as a boy he first learned to pitch a ball.

One of the most common examples of an act which is even more involuntary is speech. Seldom are we conscious of the movements of our vocal organs while talking, nor are we often conscious of how we are going to move them before we start. Our minds for the most part are occupied with the direction of the thought, the effect we are making, and to some slight degree with the sound of our voices. That we are usually unconscious of the way we talk is made clear by contrast when we hesitate before a foreign word which is difficult to pronounce. One is even more unconscious of his actual vocal movements when he reads an uninteresting book aloud to a friend, while he thinks at the same time of something quite different. He is almost unaware, not only of his motor response, but even of the meaning of what he is reading. Through long training the words on the page (the visual stimuli) set off the appropriate action patterns and he reads quite intelligently with a minimum of attention—a minimum of intention in fact.

A day is replete with such semi-voluntary acts, acts that hardly touch the conscious level. Even while one writes with concentration at his desk these incidental actions intrude and pass. One looks up for an instant while pondering a difficult problem and his eyes rest upon his pipe. His hand reaches out for it, puts it in his mouth, and the writing goes on with scarcely a break.

Automatic acts. It is possible in situations like the preceding for one to be totally unconscious of what he has done. An act in which the person is unaware of his performance is called *automatic*. Automatic acts, like semi-conscious ones, are very common. We curl a strand of hair, bite our pencil tip, tap on the floor, twist a button, entirely unaware that we are doing anything. We engage in animated conversation with a friend while walking, entirely uncon-

scious of the action of our legs. Automatic acts can also be as complex and can involve as highly an integrated set of reactions as any fully voluntary response. This fact is well demonstrated in the phenomenon of automatic writing, when a person writes the answers to questions put to him without the slightest idea of what he has written. It seems evident in this case that his hand has been guided by subconscious processes. Thus the method is often used to find out what lies below the level of his consciously controlled behavior.

The examples which we have examined in these last sections illustrate the various forms of action, from the fully voluntary through the different degrees of semi-voluntary behavior to unconscious automatic acts. Such a classification, however, is by no means clear cut. Although we can have acts that are entirely automatic and unconscious, almost all voluntary acts contain some automatic process. In fact, such acts as piano-playing, when performed by a proficient person, contain so much automatic response that it is customary to use the word automatic rather than voluntary in regard to them. Here action has become so well established a habit that correct response follows immediately upon stimulus, whether the musician is using the score or playing from memory, that is to say, whether the stimuli are the musical notations and the preceding finger movements, or the latter alone. It is, indeed, frequently difficult to say whether an action is entirely automatic or not, as when the musician plays softly over the keys while conversing with a friend, or when a telegraph operator taps SOS on the desk with his finger while he is reading an engrossing detective story. The important point is that most of our responses are a mixture of the two types, automatic and voluntary.

Voluntary control of movement. What do we have to do in order voluntarily to make a response? It was at one time supposed that, if we could call to mind how the muscles would feel when moved in a certain way, that is to say, if we had a clear memory of the kinesthetic sensation produced by

the movement, we could then move those muscles appropriately. It was even sometimes supposed that such a memory of a movement must necessarily precede the movement which we desire to make. Research, however, has shown that kinesthesis alone or even when combined with a visual memory of the movement is not a sufficient preliminary process to produce at will a movement that had never before been voluntarily initiated.

In certain experiments, persons, who could not move their ears voluntarily, had the ear muscles stimulated electrically so as to produce the movement. These persons felt the movement and saw it in a mirror; still they could not move their ears voluntarily. In attempting to move them, they had the same sense of helplessness that they experienced before the electrical stimulation. In this attempt, however, they moved other muscles—those of the brow, jaw and cheek. In this way the ear muscles were at last brought into the reaction pattern, but it was only then that the appropriate kinesthesis helped to develop full voluntary control of the ears (2).

The results of these experiments bring out the important point that *kinesthesis* aids voluntary control only *if it accompanies a movement actually initiated by the subject* and thus becomes a conditioned stimulus for that movement. Kinesthesis of passive movement alone, such as is initiated by an electric shock, is not a sufficient antecedent for voluntary control, since the motor nerves to the muscles in question have not in this instance been innervated. These facts also give us a picture of the genesis of all voluntary movement.

There is a method of re-educating paralytics which throws additional light upon the subject of muscular control. The paralysis in question is usually caused by injury to the cerebral motor area concerned with the particular movements. The person may strongly will to move the limb, but no movement results; nor does it usually help him to learn if someone moves the limb for him, any more than the electric shock is an aid in initiating the movements of the ear. When, however, he makes a great effort he may move unimpaired muscles, and that

movement will involve the paralyzed muscles. Gradually in this way, probably because of the formation of new nervous connections, he slowly learns to make the desired movement. Recovery may be hastened by having the patient play athletic games, since he is thus placed in a situation in which he is likely to make strong muscular efforts.

It is clear from the above descriptions that the *first movements* of our muscle groups are *unconscious* and *involuntary* and that later, in the integration of these movements into complex patterns of behavior after they have been 'accidentally' innervated, they come under conscious voluntary control. The fact that conscious habits tend to develop into automatic action has given rise to the opposite notion that all movement is at first conscious, but this view is no longer tenable.

Ideo-motor action. Although one often calls to mind kinesthetic images of the intended movement before making the response, especially when learning difficult movements, such imagery is not necessary as a preliminary to a voluntary act. The idea may be merely the verbal instruction to move in such and such a way (see p. 454). In fact it may be any kind of imagery even remotely connected with the act to be performed. The idea of the action itself, however, whatever may be the imaginal terms in which it is carried, seems to be the most compelling antecedent to action, for there is under such circumstances a very strong ideo-motor connection, as is plainly shown in the following experiment.

If a recording instrument is placed on a person's head, so that a graphic record of his head movements can be got, it is found that when, with his eyes closed, he merely thinks of his head moving to the right, the record shows that his head has actually made a slight movement in that direction. When he thinks of moving his head to the left, the record indicates that such a movement has been made. However, the person himself is unlikely at any time during the experiment to realize that he has made an actual movement.

The feat of muscle reading, a form of 'mind reading,' is

based on this fact of action following the thought of it. The performer takes the hand of an individual and tells him to think as hard as possible of what he wants him to do. If the individual wishes him to go toward the window, his hand will make a slight movement in that direction, which the performer, who is supersensitive to such weak muscular responses, will immediately feel and use as a cue. Animals are particularly acute in detecting these unconscious movements. A trained dog may be able to pick out the correct one of a series of playing cards spread on the floor if persons who know the correct card are nearby. In thinking of the correct card, the onlookers unconsciously turn their eyes for a fraction of a second, and perhaps their heads also, toward the card in question, and the dog notices this direction of regard.

The same phenomenon is also illustrated in persons who dislike high places. The idea of jumping comes so strongly to mind that they fear it will go over into action. Nearly everyone has had at some time so vivid an idea of the act of jumping out of the window at which he was standing that he has wished to withdraw from the window in order to avoid the danger.

Empathy. Innumerable examples of a similar nature could be taken from daily life, for ideo-motor action is a very common experience. Still more frequent, however, are the incipient movements, sometimes too slight to be readily detected, at other times quite noticeable, which are aroused in us by the movements in our environment. An obvious example may be observed at a football game where one's own team, let us say, is holding on the one-yard line. Then one may push actively with the players until he suddenly realizes that he is actually pushing his neighbor. Or again, when spectators watch an acrobat climb to the top of a pole balanced on the head of a colleague and swing back and forth with the tottering pole, the whole crowd sways in unison.

This sort of movement occurs also in looking at statues, buildings or pictures, or in listening to music. We may feel the

thrust of the foot or the tension of the outstretched hand of the statue, the weight of the arch on its column or the rise of the column itself, and the direction of the lines and weight of the represented mass in the picture. When we listen to music, we often find ourselves following the rhythm with some part of our body. Even the rhythmical click of the car wheel over the rail may arouse a motor response. For the most part, we are not conscious of these movements in ourselves, since we are occupied with the perception of the object. Nevertheless our responses, though unconscious as such, give dynamic quality to these perceptions. The lines of the picture become lines of force, the represented mass has weight, the rhythm of the music seems to flow smoothly, the curves of the architecture appear to have the grace of a moving object.

Gracefulness is indeed a very interesting example of the manner in which our movements influence our perceptions. Herbert Spencer rightly defined gracefulness as the characteristic of movement which is produced with the minimum amount of effort. The accomplished ballet dancer tries to conceal any movements of face or body which would suggest strain or effort to the observer. The more successful the performer in achieving this end, the more graceful the dance appears. Her face is usually placid or even smiling; she whirls and springs with apparent ease. Little feeling of strain is aroused in us as we watch, and she therefore seems to move in the air with the lightness of a leaf blown by the wind.

Now we do not think of these responses, but rather of the characteristics they produce for us in the objective world. It is as if we had projected our own experience into the object of our perception. This sort of projection has been termed *empathy*, or a feeling of ourselves into the object of regard. Empathy is an important factor in much of our esthetic experience, which is that side of our life concerned with the beauty of things.

Suggestion. In the broad sense of the term, suggestion plays a large role in our lives of action. The immediate per-

ception of an object most frequently leads to some response which depends upon previous experience with the object. The flame suggests withdrawal of the hand. The sight of a half-read book suggests continuing the story; without any intervening thought the student picks it up, when he had fully intended to settle down to study. The sleight-of-hand performer, by a movement of the other hand, suggests a shift of the attention of the audience away from the hand that is doing the trick. In the empathic perception of lines and mass there is the direct suggestion of some motor response. In these instances the term suggestion is used legitimately; however, it is usually restricted to that action which is brought about by a verbal instruction. One acts through suggestion when he responds to the written or spoken word *uncritically*. In most instances such a response is immediate, but it may on occasions be delayed.

A person is said to be highly suggestible when he lacks firm convictions of his own. Most of us can act through suggestion a thousand times a day without losing individuality, but there is the extreme case when a person has so few firm convictions of his own that no counter argument enters his mind when he is presented with an important course of action. Conversely, there is the negatively suggestible person. He almost invariably has some reason for not doing what is desired of him. The first type cuts out the coupon of the advertisement at once and mails it. The second throws it immediately in the waste-paper basket. It is thus that attitudes toward suggestion determine action in the large as well as the small affairs of life. Degree of suggestibility is an essential feature of the individual personality.

Hypnotism. The hypnotic trance is an extreme state of suggestibility, which may be induced in varying degrees in most normal persons who are willing to cooperate with the hypnotist. The person who has been hypnotized is in a condition resembling sleep, except that he can respond adequately to external stimulation when the hypnotist suggests it. As the

subject's mind is free from the ordinary inhibitions and resistance, he readily carries out the instructions given him by the hypnotist, provided that the task does not conflict with his more fundamental convictions. He will commit an artificially arranged crime but, contrary to popular belief, he cannot be induced to commit an offense which really contravenes strong tendencies of ethical conduct. Under hypnotism only the more superficial inhibitions are removed.

It has been supposed that under hypnosis a person's senses are keener and his strength greater than normal. Experiments, however, have shown that this is not the case. There is little if any difference in his threshold of sensitivity, and the feats of strength he performs under hypnosis he can also do in his normal state if he is willing to make great effort. An interesting phenomenon of the hypnotic state is *post-hypnotic suggestion* which is an extreme form of delayed response and shows clearly the working of the determining tendency or mental set. For example, if a person under hypnosis is told by the hypnotizer to drink a glass of water at a specified time after coming out of the hypnosis, he will accurately obey the instruction. If asked why he took a drink, he will probably reply that he was thirsty, since he will have no recollection of the real cause of his action (see also p. 472).

QUANTITATIVE DESCRIPTION OF ACTION

Reaction time. In the preceding sections a number of qualitative aspects of volitional acts have been described. It is now necessary to consider the speed of reaction and the conditions which determine the speed. The problem of the reaction time arose in connection with an incident that occurred in 1796. A certain astronomer at the Greenwich Observatory in England dismissed his assistant because the latter's observations of the time at which stars cross a cross-hair in the field of the telescope were almost a second later than his own. Twenty years later it was discovered by checking the observations of different astronomers that the discrepancies were due

to more fundamental differences in the manner of reaction than would be produced by mere carelessness. The conclusion was reached that such measurements, depending upon speed of reaction, were affected by what was then called the personal equation, that is to say, fixed individual differences in reaction time.

One of the most accurate arrangements for the measurement of human responses is illustrated in Fig. 128. Its main feature

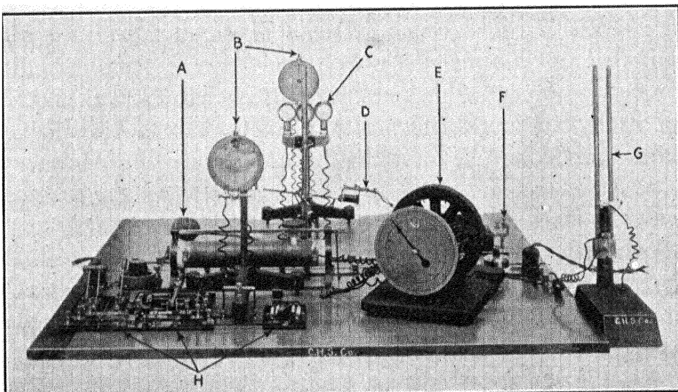


FIG. 128. INSTRUMENT FOR TIMING REACTIONS.

A, bulb for response key; *B*, voice keys; *C*, light stimulus; *D*, relay for touch stimulus; *E*, chronoscope; *F*, relay for sound stimulus; *G*, tuning-fork for time control. Courtesy of the C. H. Stoelting Company of Chicago.

is a chronoscope or timing device, which consists of a synchronous motor and a dial whose hand is attached to a magnetic clutch. Two telegraph keys are wired to the instrument in such a way that when one key is pressed the clutch engages with the motor and when the other key is pressed the motor is released. The subject is seated at one key and the experimenter at the other and the motor is started. The experimenter gives a preparatory signal and a second or so later presses his key. The circuit is thus closed, the stimulus (sound, sight, touch) is given the subject, the clutch is engaged and the hand on the dial revolves. The subject presses his key as quickly as possible upon perceiving the appropriate signal,

thereby breaking the circuit and releasing the clutch and the hand stops. The revolutions of the hand are recorded on the dial, and as the speed of revolution is already known, the time that elapsed between the pressing of the two keys, that is to say between stimulus and response, may be read from the dials in milliseconds.

Various stimuli and types of response may be used. For example, the experimenter may signal by means of a clicking sound produced by a relay, or he may give a tactual stimulus by means of an automatic contrivance that presses on the subject's hand. Flashing of an electric lamp may be used as a visual stimulus, and, if a 'choice reaction' is desired, the experimenter can illuminate in haphazard order either a green or red lamp. For word reactions one uses a voice-key containing a thin diaphragm which vibrates when spoken against, thus temporarily breaking the electric circuit. The experimenter may speak into one voice-key starting the clock and the subject may speak into the other key stopping it. There are other possible arrangements, but in all the clock is started and stopped automatically and, if the alternating current is well regulated, reaction times are obtained which are accurate within a few milliseconds.

Simple reactions. In the *simple* reaction experiment, the subject is generally instructed to respond by pressing a telegraph key as quickly as possible after the signal is given by the experimenter. It is found that individuals vary among themselves in speed of reaction, and also that the reaction time of the same individual varies according to the *sense organ stimulated*. The following table shows the approximate range of the reaction times in milliseconds for the different senses:

Kind of Stimulation	Reaction Times
Visual.	150 ms.–225 ms.
Auditory.	120 ms.–185 ms.
Tactual	115 ms.–190 ms.
Olfactory	200 ms.–800 ms.
Gustatory.	305 ms.–1080 ms.

Kind of Stimulation	Reaction Times
Pain	400 ms.–1000 ms.
Cold	150 ms.
Warm	180 ms.

The reaction times to painful stimuli are especially long. There is a considerable latency between the application of a stimulus and the consciousness of pain, and it is due in part to this fact that the reaction times are so long. The reaction times for warmth and cold vary according to the manner of application of the stimuli. The reaction to taste varies with the part of the tongue stimulated and the kind of stimulus. The reaction is shortest for salt and longest for bitter. The reaction time for touch varies according to the part of the body stimulated and to the limb making the response. The reaction time to a stimulus applied to the forehead is longer than to one applied to the hand. The reaction of the right hand to a tactual stimulus applied to the same hand is shorter than to a stimulus applied to the opposite hand. The reaction to light is faster when the light falls on the fovea than when it falls on an eccentric position on the retina, and the time increases continuously with the distance of stimulation from the fovea. Reaction is more rapid to binocular than to monocular stimulation.

The *duration* of the stimulus also seems to affect the speed of reaction. There seems to be an optimal duration which varies with the subject. If the duration of the stimulus is increased or diminished from this optimal value, reaction times are longer.

In the case of vision, the form and the area of the stimuli influence the rate of the reaction time.

In most experiments upon reaction time, a preparatory signal is given before the presentation of the stimulus. It is found that the reaction time varies with the *length of the interval* between the preparatory signal and reaction signal. Constant intervals between 2 and 4 sec. give the shortest times. If the preparatory signal is varied within a series so that the subject never knows exactly how long he has to wait for the reaction

signal, the optimal interval ranges between 12 and 16 sec. The act of preparation seems to be the chief factor involved in these results. If the interval is too short, the subject has not sufficient time to 'get set' and the stimulus may come before he is quite ready. If he has to wait more than 4 sec., the interval becomes monotonous and he cannot hold his attention entirely on the task. If the interval is varied, he is unable to assume a constant preliminary set or attitude of expectancy and thus requires a longer interval for his quickest reaction.

Distraction usually lengthens reaction time, but sometimes it decreases it, and occasionally after a short initial disturbance it leaves the times unaffected. The paradoxical effect of distraction has been found in other experiments where concentration is necessary, and it is explained by the fact that some persons use more effort to concentrate when there is an obstacle to overcome. City dwellers become so accustomed to concentration 'in spite of' the noise of the street that they often have difficulty at first in working efficiently when they go into the country.

Reaction times to all kinds of stimuli decrease with an increase in the *intensity* of the stimulus. This decrease in time is most marked in the range of weak intensities.

Although an individual's reaction time varies according to the nature of the stimulus, the question arises whether, if he is quicker than his fellows in his response to visual stimuli, he will also respond more quickly to auditory and tactual stimuli. In other words is there a speed characteristic of response that runs through all of a person's motor reactions? In a series of experiments it was found that the average intercorrelation of simple visual, auditory and tactual reaction times was 0.86 in one experiment and 0.72 in a second experiment. These results indicate that if a person excels in speed of reaction to one kind of stimulus, there is a good chance that he will excel in other kinds also (9).

Sensorial and muscular reactions. If one runner starts at the crack of the pistol sooner than his rival, it is due to a

difference not only in respect to the neuro-muscular system but also to the attitude of the contestants. It has been shown in the laboratory that there are two types of reaction, *sensorial* and *muscular*. In the sensorial type, the subject's attention is directed to the stimulus by the instruction, and in the muscular type to the response which he is to make. In the extreme form of sensorial reaction, there is a clear anticipation of the coming stimulus, often with a steady fixation in the direction of its appearance. In the extreme muscular reaction the idea of movement in terms of kinesthesis is dominant. If the subject is allowed to react 'naturally,' there is usually an attitude midway between these two forms, or an alternation of the two.

These differences in attitude cause differences in reaction time. When the reaction tends toward the sensorial type, the time is longer than when it tends toward the muscular type. In the table on reaction times, page 447, visual reactions range from 150 ms. to 225 ms. It is probable that the 150 ms. was obtained under a motor set and the 225 ms. under a sensory attitude. With practice one tends to become increasingly motor until the reaction becomes practically automatic; then the finger movements occur with little conscious intention as soon as the signal is given. With this extreme motor attitude, however, premature reactions are not infrequent, as one finds not only in the laboratory but also in such situations as racing. A runner who is of the extreme motor type often makes a false start. Some runners, however, prefer to be sure of the signal, even though they are a little late; and so it is with people in general. There are those who are slow, safe and sure, and those who go off 'half-cocked.'

Discrimination and choice reactions. Most of our reactions in life are not as simple as those just described. It is seldom that we are so sure of what is going to happen that we can be definitely prepared to react automatically at maximum speed. Almost always at least some uncertainty exists as to what is going to occur. For instance, if the runner is not alert

he may start at the sound of an auto back-firing instead of at the pistol shot. Some *discrimination* is generally necessary for a correct response.

In the laboratory this ambiguous situation is produced by varying the stimuli. For example, the subject may be instructed to react only to a red light when both red and white signals are used in haphazard order. The necessity of recognizing the correct signal increases the average reaction time above the time of the simple reaction, and, the more muscular the set of the subject is, the more likely he is to react to the wrong light. This situation may be further complicated by requiring a *choice* between two or more reactions as well as a discrimination between stimuli. The subject may, for example, be instructed to respond with the right hand if the light is red and with the left hand if the light is green, or with the right hand if the red light appears on the right of the green and with the left hand if the stimuli are reversed. These further complications increase the reaction time.

In regard to discrimination, it is found that, the more the stimuli resemble each other, the longer are the reaction times. If black and white are used as stimuli the reactions are quickest. Red and green come next, red and blue next, then red and yellow, and the slowest times occur with red and orange. If tones are used, the reactions to tones differing by 16 cycles is quicker than to tones differing by 12 cycles, and much quicker than to tones only 4 cycles apart. When lines differing in length are the stimuli, the reactions are longer the less the difference between the lines. The reaction time is, for example, shorter for discrimination between lines of 10 mm. and 13 mm. than for 10 mm. and 12 mm. The lengthening of the time in these experiments is due to the increased difficulty of discrimination and also the increased difficulty of the subject's assuming an adequate preliminary adjustment for the movement.

Word reactions. The most usual reactions in life are *verbal*. Numerous experiments have been devised to determine the nature and speed of such responses. The usual method is

to present a word visually or vocally with instruction to respond with the word that is suggested by the stimulus word. The time, which may be taken by a stopwatch or by means of voice-keys and a chronoscope (see pp. 446f.), indicates the speed of the association of ideas for the individual tested. If the subject is told to respond with the first word that occurs to him, the association is termed 'free.' Frequently, however, the instructions are more limited. For example, a general term indicating a class, such as "animal," is given and the subject is required to reply with the name of a member of this class, such as "bear," or he is instructed to respond with a word opposite to the stimulus word, namely with an antonym. Many other variations in instruction may be given. These associations are partially determined from the start and are called controlled associations. Experiments of this nature have been extensively employed in investigations of the nature of the thought process.

Practical use of reaction experiments. An individual's ability in practical affairs depends in part upon his speed of reaction. It is therefore frequently of value to know how rapidly he reacts in a given situation and under varying conditions, and how he compares with other individuals under similar circumstances. It is also of interest to know how much he may improve his speed and accuracy by practice and under the incentive of increased interest in the task.

Reaction time is often an important factor in vocational selection as well as in determining the individual's aptitudes as a basis for vocational advice. For example, in the selection of telephone switchboard operators it is obvious that speed of response and relative freedom from errors are essential requirements. A consideration of the same characteristics is necessary in the selection of chauffeurs and machine operators. According to tests made on taxi drivers, those men with the greatest number of accidents have the slowest reaction times. Those who have the fastest reaction times have also many accidents, perhaps because they are overconfident and take

chances. It is therefore desirable to select drivers from the group whose reaction times are neither very fast nor very slow.

The association-reaction experiments have been used with some success to determine guilt. Words are selected which are related to the crime and these are interspersed with 'neutral' words. The words of this combined list are read to the subject, who must answer as rapidly as possible to each one with any word he can think of. Anyone knows from his own experience that when he is faced with an embarrassing situation—one that is emotionally toned—he is apt to hesitate and often to reply foolishly or irrelevantly. In the 'crime' experiment there is exactly such an embarrassing situation for the guilty person. Therefore the tendency is for the reaction time to the relevant words to be unusually long, or at least to vary more than the reaction times to the neutral words, and in addition the quality of the words is often different in the two cases.

This same method is used in order to discover suppressed complexes—the memory of painful experiences that are held in a subconscious state and that often give rise to abnormal mental conditions. Owing to the fact that such complexes are, like the concealed knowledge of the guilty subject, highly emotional in nature, the two test situations are very similar.

THE NATURE OF THE WILL

The will. A person decides to go to town. He walks down the stairs, puts on his coat and hat, opens the door, gets into his car and starts the engine. Common sense says that he has willed to do these various acts. Or again one is trying to read a difficult passage in a textbook. His mind continues to wander from the book to irrelevant matters, until finally with great effort he succeeds in concentrating on the work at hand. It is usual to say that he has had to use his will power. No fault can be found with such an expression of ordinary speech, but the psychologist desires to know what is this conscious experience, and what is the general process that one calls 'will.'

We have seen that in voluntary action there is frequently

some idea of the movement or thought of the instruction preliminary to the movement itself. Many such descriptions of the period antecedent to action have been obtained, principally from the experiments on reaction time. If the response is to be made with the right hand to a red light and with the left hand to a white light, there may be, for example, some verbal imagery of the instruction with perhaps tension in the arms and a visual association of right hand with red light, etc. One can say that the task or problem is represented mentally in the above manner. This attitude of the subject is called the *set* toward the task, and as the experimental series progresses the set becomes increasingly muscular in nature, so that eventually the movement occurs immediately upon the appearance of the stimulus without any intervening mental state. This set may be either positive or negative. In the experiment with the red and white lights, the set for the right hand is positive toward the red light and negative or inhibitory toward the white light.

An experiment can be arranged to investigate the muscular set by placing a rubber bulb or tambour on the reaction key in order to measure the amount of pressure of the finger. It is found that the finger frequently makes an actual anticipatory movement of downward pressure on the key before the real movement is carried out. A good example of muscular set is that of the football player who has in mind just what to do in answer to the play of his opponents. Offside play is frequently due to an over-intensified set similar to the set in the experiment just described, when the finger made an anticipatory movement. In another experiment it has been shown that there is greater effort as recorded in preliminary tension when the task is difficult than when it is easy. In the daily acts such as those cited in the beginning of this section, the sequence of events follows so rapidly and one is so little given to introspection on how one does a thing, that the essential attitudes escape attention. They can be observed, however, especially if the task is difficult.

The second phase of the voluntary act is its execution. Here

the set or attitude carries through as a directive mechanism called the *determining tendency.* (For its description in relation to thought see pp. 469-472.) It is this directive tendency which makes an act more than a rigid mechanical sequence of events, such as is found in the movements of a machine. The determining tendency, formed by the preliminary set, causes one to respond correctly with the right hand instead of the left on the appearance of a red light, or it is responsible for the selection of the correct words of a sentence.

In a voluntary act there is no special force that one can call the 'will.' What one feels in an experience of 'will power' is the muscular tension involved, such a tension as that of the muscles of the forehead when one wrinkles his brow in an effort to concentrate. It has been argued that, since a person paralyzed in one leg experiences an effort of will when he tries to move the inert limb and yet does not move it, the will experience obviously cannot come from these muscles. But what actually happens is that unknowingly he moves the other limbs or some other member, and it is this movement that gives him the impression of will power. The will, then, so far as experience is concerned, turns out to be the preliminary attitude and the experience of movement with the knowledge that the movement follows directly on the attitude and has not been caused by any external force. We *know* that we have made the movement. It is unfortunate that 'will' is a noun, for it implies some agent, faculty or special kind of energy. *There is 'willing' but not a 'will.' Willing is a process which one calls a voluntary act.*

Abulia. Everyone at times finds it hard to get down to work, especially if the task is difficult or if there has been a long interruption such as a summer vacation. He invents all sorts of excuses to postpone action. This state is a mild form of what in its more pronounced manifestation is called *abulia,* an inability to arrive at decisions. A person in this condition is not inactive or lazy, for he will perform useless acts any number of times, such as going down to the cellar every few min-

utes to see if he closed the furnace door or repeatedly washing his hands or cleaning his glasses. He may get up in the morning full of energy and resolution only to find himself caught in the toils of petty activity directly he has finished his breakfast. This state of vacillation and procrastination gives him a sense of inferiority which aggravates the situation. Authorities believe that abulia is due either to lowered vitality or to repressed unconscious wishes, which, being impossible of fulfillment, under existing conditions, cause fear and anxiety with an accompanying tendency to retreat from reality.

Motives. The fundamental motives for action are the internal stimulations that are generally termed *animal drives.* Chief among them are hunger, thirst and sex. They are among the annoyers which in the early stages of development cause those restless, random movements previously referred to. The feeling of hunger is produced by contractions of the stomach. which in turn are caused primarily by the chemical condition of the organism. Thirst is caused by a dryness of the mucous membrane of the mouth, which in turn results from lack of water in the system. These are both annoying experiences, inducing a restlessness which continues until the unpleasant conditions disappear—in the case of the infant until it either finds milk or is helped to it. From what we have already learned, it is obvious that with experience the sight of the nipple or the bottle will arouse the appropriate action, and thus the food becomes the goal of its action, just as the food at the end of the maze becomes the goal for the rat's activity and the motive for its behavior.

As the infant matures, there soon ceases to be this simple correspondence between hunger and food, and thirst and drink. Other habits come in to complicate the situation. The child will eat when it is not hungry and will sometimes through stubbornness refuse food when it is hungry. The main reason, however, why the primitive stimulus-response relation of hunger and eating is not more frequently seen is that, in the well-

ordered routine of civilized life, we are seldom hungry in the sense of having these special stomach contractions.

Some idea of the factors that complicate the feeding response may be had from the experiments on chickens. When a dish of corn is placed before a chicken it eats a certain amount and stops. If, however, another dish containing a different kind of food is placed before it, the chicken eats again as if it had not previously eaten. By changing the kind of food in successive offerings, the chicken can be induced to eat an enormous amount of food. The eating response continues long after the disappearance of the hunger drive. It is also found that, when another chicken is brought in to eat from the same dish, the food consumption by the first chicken is greatly increased, a phenomenon observed with other animals and also with children.

Numerous studies have been made on the sex activities of animals, especially those of the female. There are two periods of unequal length in the estral cycle of the female, a short period when the ovum comes from the ovary to the womb for fertilization and a longer period between the successive shorter ones. During the shorter period the female is much more active than at other times, and this increased activity shows itself in two ways. There is a heightened general activity and a specific sex activity. In order to test the general activity, a female rat was placed in a cage equipped with a wheel like those provided for the exercise of squirrels. An automatic counter recorded the number of revolutions of the wheel. It was found that the female was much more active when she was in heat than during the intervening period.

For the purpose of testing specific sex activity, an obstruction box has been used, a diagram of which is shown in Fig. 129 (13). The female rat is placed in compartment *A*. In order to reach compartment *C*, she must pass through the alley *B*. The floor of this section is covered with an electric grid, which enables the experimenter to give the animal a weak shock. If she crosses the grid, she steps on *E*, which releases a door d_2 and allows her to reach the male in compartment *D*. It was

found that when she was in heat, she would cross the charged grid frequently and with little hesitation, but that at other times she would scarcely ever cross it. There are no cycles of sex behavior in the male rat, but he crossed the grid more frequently during a period of sex deprivation than after he had been satiated.

The same method is used to test the strength of other drives. When food is the incentive, the animal is first given a training period to get it acquainted with the situation and then a twenty-minute test is made with the grill charged. During the

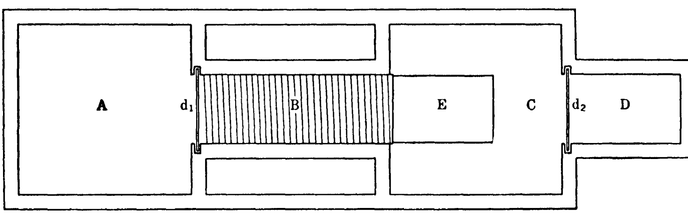


FIG. 129 DIAGRAM OF FLOOR PLAN OF OBSTRUCTION BOX.

A, entrance compartment; *B*, obstruction compartment (electric grid); *C*, *D*, divided incentive compartment, *E*, release plate, d_1, manually operated door between entrance compartment (*A*) and grid (*B*); d_2, automatic door operated by animal's stepping on release plate (*E*). From Jenkins, T. N., Warner, L. H., and Warden, C. J., Standard apparatus for the study of animal motivation. *J. Comp Psychol.*, 1926, 6, p. 366

experiment the animal is deprived of food except for a nibble it obtains when it reaches the food box. It was found that the male rat would cross an electric obstruction least frequently directly after having been fed. His tendency to cross an obstruction increased with an increase in the length of the starvation period up to four days, and from this point on it decreased. The female rat showed a maximum tendency to cross much sooner than did the male. This apparent decrease in the hunger drive after a certain period is also found in human beings. Persons who have fasted for a long period of time report that the desire for food disappears after about the third day.

Experiments have also been made to test the effect of hunger on more complicated activities like maze learning.

In one investigation three sets of rats were used. One was kept in a half-starved condition by reducing the ration of food; the second group was given a food that did not contain the proper amount of nourishment, and the third group was fed normally. The results showed that the undernourished rats surpassed the well-nourished rats in learning the maze.

The obstruction method has been used to test various other animal drives. A comparison of the effects of the different drives indicates that under given conditions the maternal drive is the strongest, thirst and hunger come next and the sex drive is the weakest of the four. This order depends, however, on the degree of deprivation of the animal and upon the particular apparatus used.

Some understanding of the potency of the fundamental organic urges of animals may be obtained from these experiments. It cannot be doubted that they are as fundamental to human behavior, but they are seldom as apparent. Even the situation of the hungry animal seeking food through the numerous alleys of a maze is less complicated than most human activity. The human being with his greater intelligence acts toward very remote goals, and his motives such as jealousy, greed, ambition and love are complex and more sustained. There are thus usually many steps between the first stimulus and the final response. A dog will directly attack a person who attempts to take his food. A starving people will start a revolution, cut off the heads of its rulers and try several different forms of government before it obtains its goal. Most of our daily behavior, simple as it may seem, is toward a distant goal which requires many separate acts to reach. If one wishes, for example, to write a letter, he must go into the study, sit at the desk, take up the pen and dip it in the ink. Each act is a necessary antecedent of the next, and the last act is linked to the first by the purpose which runs through the series. This purpose is a goal-seeking which is represented functionally by that set or determining tendency previously described. Its physiological nature is not as yet fully understood, but it controls and gives general direction to behavior.

EXPRESSIVE MOVEMENTS

It is generally supposed that an individual is self-consistent in his personality, and that this self-consistency appears in his activities in the sense that his style of writing has factors in common with his manner of speech, his facial movements with his posture, and his gestures with his manner of walking, etc. Involved in this belief is the assumption that there is a self-consistency in each of the expressive acts, so that an expressive act performed by an individual today will be similar to that performed tomorrow.

Common belief usually contains at least an element of truth, but it is generally necessary for the psychologist to test popular assumptions. An extensive study of this problem of expressive movements has yielded some interesting results. A number of activities were quantitatively examined, the chief of which were reading and counting, walking and strolling, the estimation of distances and angles and of weight, the strength of grip as it appears in activities like the handshake, and tapping with the finger, with the hand and with the leg. The measurements show that there is a certain constancy in the repeated performance of a test. When the test was repeated on the same day the average reliability is +0.75, and when given at different sessions, +0.64. These figures mean that a person who has a certain rank in a test is likely to maintain that rank when the test is given again. It would seem then that the characteristics of an individual's performance of a certain task are specific to that individual and may be expected to appear whenever the individual performs the task. It was also found that, if a person obtained a certain score in a task like tapping with the finger, he would probably continue to make approximately this score even if he changed from a finger of the right hand to that of the left or to the foot. Not only was this transfer true for the same task, but even different tasks performed by different groups of muscles showed agreement.

Such results indicate that the individual possesses some gen-

eral psychomotor traits. The speed of performance is not consistent throughout all the tests, but an individual seems to have (1) a verbal speed which is common for activities such as reading, counting and some parts of handwriting and blackboard writing; (2) a drawing speed for hand drawing on paper, for foot drawing in sand and for some features of blackboard drawing, and (3) a rhythm speed for finger, hand and leg tapping and for pressure. In short, there was no evidence of a general personal tempo, but there were strong indications of three group factors of speed (1).

One may draw the general conclusion from these studies that, although no one central factor of the organism controls and shapes all our acts in a characteristic manner, neither is each act an isolated and specific affair. Evidently general attitudes of persistent motor dispositions influence diversified types of movement. Certain characteristics of the individual, for example, may be inferred from his handwriting. His handshake and his wave of greeting may give one a certain insight into his general disposition, and even the manner in which he holds a pencil or a cigarette is frequently enlightening. These movements and gestures are expressive of the thought and the personality of the individual.

REFERENCES

1. Allport, G. W., and Vernon, P. E. *Studies in expressive movement.* New York: Macmillan Co., 1933.
2. Bair, J. H. Development of voluntary control. *Psychol. Rev.,* 1901, 8, 474-510.
3. Carmichael, L. Origin and prenatal growth of behavior. In *A handbook of child psychology.* Second edition. Worcester, Mass.: Clark University Press, 1933. Chapter 2.
4. Carmichael, L. The development of behavior in vertebrates experimentally removed from the influence of external stimulation. *Psychol. Rev.,* 1926, 33, 51-58.
5. Coghill, G. E. *Anatomy and the problem of behavior.* New York: Macmillan Co., 1929.
6. Garrett, H. *Great experiments in psychology.* New York: D. Appleton-Century Co., 1930. Chapter 9.

7. Holt, E. B. *Animal drive and the learning process.* New York: Henry Holt and Co., 1931.
8. Hudgins, C. V. Conditioning and the voluntary control of the pupillary light reflex. *J. Gen Psychol.,* 1933, 8, 3-51.
9. Lanier, L. H. The interrelations of speed of reaction measurements. *J. Exper. Psychol ,* 1934, 17, 371-399.
10. Peak, Helen. Modification of the lid-reflex by voluntarily induced sets. *Psychol. Monog.,* 1931, 42, No. 188.
11. Peak, Helen. Reflex and voluntary reactions of the eyelid. *J. Gen. Psychol.,* 1933, 8, 130-156.
12. Travis, L. E., and Patterson, M. Rate and direction of the contraction wave in muscle during voluntary and reflex movement. *J. Exper. Psychol ,* 1933, 16, 208-220.
13. Warden, C. J. *Animal motivation.* New York: Columbia University Press, 1931.

CHAPTER 18

THOUGHT

Two boys on the upper platform of a Roman tower were overheard discussing its height. The boys walked round the platform, leaned over the rail, looked at the ground. Something like the following conversation took place:

"How high do you think it is?"

"It looks about seventy feet."

"How do you know?"

The first speaker looked over the side again. The other answered, "We're more than twice as far up as that house. That's about thirty feet."

The first replied, "You can't tell from that. Let's count the steps or something."

There was a short silence after which the first boy cried, with something like a shout of triumph, "Let's drop something and time it."

A coin was produced and dropped, and the fall timed by a watch. Apparently the fall took about two and a half seconds. The sum was worked out with a pen and paper, and the announcement made, with considerable satisfaction, that the height was a hundred feet.[1]

A number of features in this conversation will prove of interest in the discussion of thought. They may be listed under eight heads.

1). There is a specific problem in the form of the question: how high is the platform?

2). There are successive steps or ideas—the first estimation

This chapter was written by George Humphrey of Queen's University.

[1] Apart from the difficulty of timing with a watch, the boys had of course neglected several sources of error, among them being their own reaction times.

of seventy feet, the questioning of this figure, the statement concerning the house, and so on.

3). These steps are not irrelevant, but are related to the problem; that is to say, they are directed towards the solution. The boys look for a coin in order to drop it.

4). The final solution is achieved by trying out first one scheme then another until a satisfactory one is found. The guess of seventy feet was rejected, and also the methods of comparison with the house and of counting the steps. The procedure is evidently one of trial and error.

5). There is much action. The boys are continually walking about the platform, bending over the edge, looking at objects on the ground. They search for the coin, the one takes his watch out of his pocket, they write their calculations, and so on. All the time they are intensely *active.*

6). In particular, action of a very specialized kind—talking—is going on nearly all the time.

7). Not only are the boys acting, but they are actively *perceiving* all the time, and their perceptions make possible the process of solving the problem. They look over the side to make the first estimation of height. They look at the house. The one feels for his watch. They *listen* for the coin to drop. They look at the hands of the watch, and so on.

8). All these actions, ideas, provisional solutions, spoken sentences and perceptions are part of a single, joint attempt to solve the problem. Beginning with the perception of the problem and ending with the announcement of the solution we have a series of activities bound together by the fact that all of them are directed towards the same end.

This example concerns two human beings, but very much the same process goes on when a single person is trying to solve a problem, except that in this case much of what is happening cannot be seen by an outsider. The individual faces the problem situation. A train of thought and action follows. 'Ideas,' action, lines of attack are tried. Some are rejected, others accepted as satisfactory. In this way a series of steps is taken which is directed towards the solution. The process involves

many different activities. Perception is active, in grasping the problem and also often during the taking of the various steps. Language is at work, mostly in the form of 'internal speech,' in which the person talks to himself rather than aloud. In addition, there is much mental imagery of other kinds, visual, auditory, and so on. *All these different activities form part of a single, unified, or,* as it is called, *integrative attack on the problem. A train of mental processes thus directed towards an end is known as directive thinking.*

Not all these features, of course, are necessarily present in every act of thought. Thus 'internal speech' may be entirely absent. A simple problem may be solved without intermediate steps. But this general plan holds good for a surprisingly large number of cases, especially where somewhat difficult problems are involved.

Stated in the simplest possible form we have then a diagram of this kind.

Problem $(\rightarrow X \rightarrow X \rightarrow X \rightarrow X \rightarrow X \rightarrow)$ Solution.

Faced with the problem, the individual initiates successive thoughts and actions, these being represented by the row of *X*'s. The intermediate processes are directed towards the solution and may be considered as parts of the total response to the problem situation.

Many points in the preceding account must now be made clearer. We shall consider them one by one.

NATURE OF THOUGHT

Integration and thought. We have said that thought is integrative, meaning that it combines many diverse data and experiences into a single line of action and thinking. Integrative action is found wherever life is found, from the single-celled amoeba up. If the organism is to live, it must meet the diversities of the outside world with a unified set of reactions. A human being has one mind, not many minds. Except possibly in abnormal cases, each of us is one personality, not many personalities. We coordinate experiences into experience.

Integration takes place on the side of thought as well as on the side of action.

The integrative action of thinking may well be seen in what is known in logic as the syllogism. Here there are two premises from which a single conclusion is reached. Thus from the premises:

> All metals are good conductors of electricity,
> Lithium is a metal,

the single conclusion follows:

> *Therefore* lithium is a good conductor of electricity.

This is then a process of obtaining one statement by combining the two preceding ones; that is to say, it is an integrative process. But, in order that this integration of thought may take place, a very complicated integration of action must first have occurred. Apart from the immensely complex processes that unify the working of the rest of the body, the very act of reading implies integration of a high order. We know from studies on eye movements that we do not spell out the words *metals*, *all*, etc., by allowing the eyes to pause at each letter, but read them as units. Reading is an integrative reaction, by which we make single responses to groups of separate letters, and at the same time experience whole words and groups of words as unities.

But more than this. Thought and action are in turn parts of the single activity of 'reading.' Indeed it is often very difficult to say where action ends and thought begins. Thinking and acting are themselves integrated, fused into a unity. The process of deducing the conclusion of the syllogism therefore presents a complex process of unification, involving integration both of thought and action. The same could easily be seen to be true of any other act of directive thinking. *Thinking directed towards an end is the conscious side of the complicated process of integration of thought and action that goes on when an individual meets and solves a problem.* When a man is solving a problem, he is forging a single line of thought-and-

action to meet a diverse situation in the world without. The various stages through which his thought passes are parts of this total integrative process.

One other point before we leave the question of integration. The experiences which are integrated in the thought processes almost always involve data which come to us through what are called the distance receptors. Thought deals to a large extent with what we see and hear, and we can see or hear a person or thing at a distance. There is great advantage in this ability to react to stimuli at a distance and to integrate the perceptions with past experiences, making a single line of action from what we have learned in the past and what we perceive at the moment. For one thing it gives the power to *anticipate*—to react to objects before they become dangerous. A world in which we could never respond to an object before it touched us would be chaos. Civilization is largely built upon our power of intelligent anticipation, which would be nearly impossible if thought had not the distance receptors at its disposal.

The problem situation. When considering the boys on the tower, we saw that the situation constituted a problem. We must now ask the meaning of this statement. What sort of a situation is a *problem* situation? The specific situation which the boys face is the spoken question—*how high is it*—taken in its total context.

Now many situations of everyday life clearly do not constitute problems at all. Such are those to which we respond by habit. It is no problem to the ordinary person to lace up his shoes or to dress himself. Yet these actions call for integrative thought and activity of a very complex kind. It is no problem for the educated adult to read his newspaper; though here again, as we have seen, integrative activity of a very complex order is employed. On the other hand, it may present a real problem for a foreigner to read the same paper. If a man knows only a few words of the language, highly complex mental activities will be involved in making out a single paragraph.

We can now see that the term *problem* is a relative one. What may be a problem for one may not be so for another. We may say that *a problem is a difficult situation, one which, for some reason or other, appreciably holds up the action of the organism, very often because of apparently incompatible elements.* When such a situation occurs it is necessary painfully to match one element of the diverse situation against another in the way called reasoning.

In a certain experiment a basket containing bananas was hung on a tree out of reach of a chimpanzee. For a while the animal moved restlessly about. Then he suddenly ran to the tree, climbed up, shook and ultimately broke the rope, and thus spilled the bananas on the ground.

Here the situation clearly contained incompatible elements, a conflict. The animal was impelled towards the bananas, but at the same time physical conditions were not such that he could obtain them directly. The result was that he remained on the spot, held by desire which he was unable to gratify. The solution came by the appropriate combination. The tree, the branch and the rope were seen as combining into a means to obtain the fruit, and action followed.

Yet it is not the situation in itself that constitutes the problem, but the animal's relation to it, the fact that it holds him up. To a man coming into the enclosure the situation would present no problem. The human organism would not be brought to a standstill by such a situation. There would be no state of perplexity, hesitation or doubt. That is to say, the human being would not have to reason in this situation, whereas many people would say that the animal did reason. Thus we reach the paradoxical conclusion that reasoning may be a mark of stupidity. The animal, perhaps, or the young child may have to reason a thing out. The adult may be gifted enough to see the solution at a glance. In the same way, one may see patients at the institutions for the feebleminded struggling heroically to thread a white bootlace in a black frame when they have been taught to thread a black lace in a white frame. The situation presents a problem to them because they

are stupid. They may be said to be reasoning on an elementary plane.

We may summarize the discussion as follows. The term *problem* is one of convenience only. It describes a situation which holds up a particular organism for an appreciable length of time. The word *reason* or *reasoning* is conveniently used to describe the conscious processes that take place while a human being (or perhaps an animal) is responding to such a situation. In practice the term is often limited to cases in which emphasis is laid on the directively ordered character of the whole process, on the original diversity of the situation or the facts of experience, and on the newness of the situation.

It is sometimes said that reasoning occurs only when a new problem is met. But every problem and every situation is a blend of new and old. Thus the banana situation presents certain new features to both man and animal, namely the position of the banana relative to the tree and the ground. On the other hand, both chimpanzee and man have seen trees and bananas many times, though the chimpanzee is more familiar with the particular tree in question. If anything, the situation is newer for the man, who does not reason.

Thus reason is not, as it was once said to be, a special faculty that distinguishes man from animals. It is not a mark of unique intelligence. One might, in fact, imagine a superior being pitying our scientists because they stupidly have to puzzle out, by mathematical reasoning, things that to a higher mind would be obvious at a glance.

THE PROCESS OF THOUGHT

The determination of thought. The stages between the problem and the solution are in general *relevant*. When the boys were trying to find the height of the platform, what they said was to the point. They did not begin a discussion on the merits of cars or the values of postage stamps, but possibilities for obtaining the height were canvassed—comparison with the house, counting the steps, using the formula. This rele-

vance of the intermediate processes to the solution gives direction or determination to thought. We have already assigned the term 'directive' to such thinking, which is thereby contrasted with so-called 'free thought' such as is found in day dreaming.

As to the fact of direction in thought, there can be no doubt. Its explanation is another matter. The problem is: what causes one mental process to follow another, relevantly to the point at issue? A schoolboy in the examination room finds printed on his paper:

$$\text{Solve the equation: } x^2 - 7x = 60. \qquad (1)$$

The boy writes something like the following:

$$x^2 - 7x - 60 = 0 \qquad (2)$$
$$(x - 12)(x + 5) = 0 \qquad (3)$$
$$x = 12, \text{ or } x = -5 \qquad (4)$$

It is assumed that this is the first time the boy has faced this particular equation, although he has had quadratic equations in his course. Thus the situation characteristically presents new and familiar elements. Since it is a new equation, its solution can hardly be said to be mere memory, as it would be if the boy had learned these four lines by heart. The boy is thinking, though in a comparatively routine way. The problem is a real one to him.

Why does line (2) follow from line (1), and so on?

The simplest theory, and one which seems to have the support of what may be called psychological common sense, refers the process to association of some kind. This may be association of mental elements, or it may involve stimuli leading to action by means of the conditioned reflex or otherwise. According to this theory, the boy's training has been planned for the very purpose of giving him the necessary habits of thought and action. Many a time at school he has said "five twelves are sixty," and "five from twelve is seven," or whatever the approved formula may be. The sight of the letters and the figures on the examination paper lets loose a chain of succes-

sive responses. All the figures are put on one side, the numbers 12 and 5 written down, and so on, by means of sheer association derived from training.

Now there is no doubt that memory is active when we are thinking, but whether it operates in such a simple manner is another question. The difficulty has always been to explain why, out of all the associations clinging to a given stimulus, the appropriate one should be released in the solution of any specific problem. According to the context the figures $\begin{array}{r}9\\\underline{5}\end{array}$ may be followed by the response 4, 14 or 45. It is not impossible to explain how this may happen by an associational scheme, but the explanation becomes somewhat complicated. Perhaps the most widely accepted alternative explanation is the one proposed by the *Würzburg psychologists.* Under the leadership of Külpe, this group undertook, in the first decade of this century, a systematic introspective examination of the whole process of thought.

The mental processes examined were of various degrees of complexity. They ranged from simple processes such as prearranged reactions to cards of various kinds or the comparison of different weights, to more complex activities such as the naming of a part of a specified object. For example, the word *man* might be given, and the word *leg* spoken in response. Very difficult problems of a verbal type were also solved, such as "Can our thought understand the nature of thought?" and "Is this true? To give each man his due would be to will justice and to achieve chaos."[2]

The work of the Würzburg school aroused the bitterest controversy on both sides of the Atlantic. It raised a number of problems which cannot here be considered. When the dust had settled, psychology found itself in possession of two new technical terms of high descriptive and explanatory value.

[2] The references to the work of the Wurzburg school will be found collected in (13). Except for this book, there is no systematic account in English. A short account is given in Boring's *A history of experimental psychology.*

These were the *Aufgabe* or task, and the *determining tendencies* which spring from it.

It was found that, when a subject is confronted with a problem, his behavior and thought are apparently determined, not only by the associations which past experience has attached to the problem, but also by a group of controls springing from the task which the subject has set himself. It is these controls that give mental life its ordered and directed character, although we are as a rule not conscious of them as such. They were called by Ach the determining tendencies.

The neatest demonstration of these determining tendencies was given in an experiment on hypnosis. A subject was hypnotized, and the suggestion was made that, after waking, he would be shown two cards with two digits on each. To the first he was to give the sum, to the second the difference of the digits. On his waking from the hypnotic state a card was shown on which were written the digits 6/2. The subject immediately said "8." When the second card was shown, with 4/2, the same subject said "2." He had no memory of the suggestion, and could give no explanation of what he had said. Nor did it occur to him that 8 was the sum of 2 and 6, or that 2 was the difference of 4 and 2.

According to the theory developed, the determining tendencies decide the course of thought by favoring certain associations that spring from the immediate situation, and inhibiting others. Thus they are to be considered as an entirely fresh principle of mental organization. They introduce directive order into the chaos of associations. They enable an answer to be given to the question: why this particular association rather than any of the many others possible? Why does the boy in solving the equation which is given on page 470 regard the numbers 60 and 7 as requiring him to find the digits 12 and 5 rather than the number 67, or the number $8\frac{4}{7}$? He was clearly unconscious of these alternatives, which, by the theory, are ruled out by the determining tendencies—the guiding, selective and inhibitory influences springing from the task which is expressed in the words, "Solve the equation."

Criticisms of the determining tendency. The theory has received wide acceptance. However, certain important criticisms of the principle have been made. Prominent are those of Selz and also those of the Gestalt school.

Selz claims that the steps of thinking follow each other by the process of completing an already established pattern of thought. Suppose that I am given the task of assigning a result to the word *cancer*. I have already in the past experienced the thought: "The result of cancer is death." The task is really equivalent to the form "The result of cancer is ——," and the completed thought follows automatically—like a reflex, he says. Selz's theory is of particular interest in view of the widespread use of what are called 'objective examinations,' in which the student is given a similar pattern to fill out. One such paper gave for completion the sentence: "The junctions between neurons are called ——." If the student had the requisite knowledge, he would automatically complete the form. According to Selz all thought proceeds on this basis. There is no need on this theory to assume that two principles are active in thinking, namely the determining tendencies and the associations on which they work. The task, namely, "find the result of," and the stimulus word, namely, "cancer," are first integrated into a single task, namely, "find the result of cancer," and the solution follows as explained.

However, difficulties arise. When the boy solves the equation, it is hard to see how there can be an adequate pattern which will cause the *correct figures* to be filled out unless the boy had solved this equation before. It was in showing how the correct figures came that the 'task' was useful. In rejecting it for the conception of a unified problem, Selz has apparently thrown out the baby with the bath water.

The Gestalt theory of thought. Again, according to Gestalt psychology Selz has not accounted for productive but only for reproductive thinking. Ready-made knowledge can explain only "stereotyped relations applied to particular conditions" (1). This is, in fact, the burden of the Gestalt criticism

of all conventional theories. The supporters of this view claim that it is impossible to do justice to the essential *newness* of many acts of reasoning if the fundamental process is held to be a stereotyped element, whether this is a habit, a conditioned reflex, a memory or a pattern. They instance the results of an experiment, for example, where the subjects were asked to devise a method of showing how much and where a metal ball was compressed when it was bounced on a hard surface. The neatest method is to paint the surface on which the ball is to be dropped, and to drop the ball while the paint is wet. Of five subjects, all reached the solution. Two required no prompting. Others had to be prompted by such hints as "think of footprints in the snow." Gestalt psychologists maintain that a 'purpose' or a 'determining tendency' has not caused the correct association to arise, for there has never been such an association in the subject's mind.

The same point comes out even more clearly in another experiment where the task was to construct, in a room, two pendulums with chalk at the end, and such that, by swinging the pendulums, two specified spots on the floor would be marked. Adequate materials were provided, and the subjects were actually shown the three necessary steps of the procedure, but not how to combine them so as to solve the problem. Even with this help, the problem was exceedingly difficult. Of 84 subjects, only 9 gave correct solutions. Only one solved it without a hint. For the other subjects the hint acted as an organizing or directing principle to put them on the right lines.

Here previous experience and habits already acquired from the past are, it is claimed, not sufficient to solve the problem. If thought is to be truly productive, experience must be made over, things must be seen in a new light and put to new uses. There must be insight, a sudden flash in which things are seen in a fresh relation to each other. The process is much like that which occurs in the perception of a hidden feature in a puzzle picture.

This view stresses the newness of thought, and contrasts productive with reproductive thinking. It is possible that experi-

ment will ultimately show the latter distinction to be one of degree only. There is an element of newness and an element of familiarity in all thought, as in all situations and responses. No adult can ever go through a process of thought that is entirely new; there are always factors contributed by past experience. On the other hand, the cases must be exceedingly rare, if they ever occur, when we *exactly* duplicate a train of thought, just as the cases must be rare when we *exactly* duplicate a meal or a walk. There are always slight differences. Thought and response contain an almost inextricable blend of new and old. By stressing the newness of the integrative processes and the fact of insight or sudden perception of new relations, the Gestalt school has, however, performed a valuable service.

The determination of thought: summary. The diagram on page 465 showed that, on meeting a problem, the organism initiates a succession of steps leading to the solution. The name *directive thinking* is given to the conscious processes occurring. The most generally accepted theory states that the various steps follow each other by the agency of memory or association, controlled by the purpose or task in hand.

Trial-and-error in thinking. Of the different steps between problem and solution, it has been seen that some are recognized to be mistakes, some accepted as correct. The boys rejected the guess of seventy feet and the methods of comparison with the height of the house and of counting the steps, but accepted as correct that of timing the stone's fall. We have already called this procedure of rejection and acceptance *trial-and-error* (see pp. 305f.).

This fact, that certain steps or solutions are accepted and others rejected, has been noted whenever the thought process has been carefully examined. Children solving problems, such as building with blocks, have been found to try out first one solution and then another. Experiments on tracing geometrical figures seen in a mirror show adult human beings feeling their way by trial-and-error until, finally, the right path is taken

quite easily. Adult subjects solving mechanical puzzles use the same process both mentally and in actual random manipulation. In fact, an elaborate examination of the method by which subjects learned to play a 'reasoning' game with matches or beads led to the conclusion that trial-and-error appears to be the invariable method in learning to solve a problem (9).

The attempt has indeed been made to *explain* all thought on this principle. According to the trial-and-error theory of thought, the animal learns by trying one action after another until the successful one is reached. In the same way, it is pointed out that human beings may actually try out responses until success is attained, although it is more usual for them to test the possibilities mentally, thus saving time and energy. In either case the fundamental fact in thinking is said to be the trial of possibilities, which are accepted or rejected, until the final solution is reached.

Let us return for a moment to the boy in the examination room (p. 470). Before finally deciding that line 3 follows line 2, more than one pair of numbers will probably be tried out. In particular, trial will probably have to be made before deciding whether line 3 shall read $(x + 12)$ $(x - 5)$, or whether the signs shall be reversed as printed. The trial may take place in the boy's head, or the figures may be actually written down.

It must be realized, however, that the statement that thinking proceeds by trial and error is a description, not an explanation. All are agreed that some solutions are rejected, others accepted. The problem still remains why this or that particular solution was tried and rejected, while another is accepted. Trial and error undoubtedly take place, and are a striking feature of all complex thinking. They constitute, however, one aspect only of the total process.

Thought and imagery. Up to the present we have concerned ourselves mainly with what may be called the mechanism of thought, namely the process by which one step follows another. We now have to consider the problem of the materials of thought. What is it that is present in consciousness

when we think? What are the items represented by the *X*'s in the diagram (p. 465)? Careful observation reveals that processes of bewildering complexity are involved, and there has been considerable disagreement about them. However, certain striking features are obvious. First of all, of course, perceptions are present, as already stated. In addition, it is generally easy to detect both images and feelings of pleasantness and unpleasantness. The latter are not usually considered to be part of the real thought process.

The imagery may be visual, auditory, kinesthetic (see pp. 347f.) or the auditory-kinesthetic imagery of words and fragments of words. Usually more than one kind of imagery is present. Kinesthetic imagery is often difficult to detect but is present in a large number of cases. When asked to multiply 12 by 13 in his head, one subject reported the visual image of a blackboard, with his own arm writing. At the same time, he had kinesthetic imagery as of the arm being moved.

Where the nature of the problem allows, successive steps will often be marked by observable action, as with the boys on the tower. On the conscious side, definite steps are often accompanied by a more or less complete word or sentence spoken internally, like the sentences in the boys' conversation. One might say to oneself: "I'll drop something." Or there may be more or less complete visual imagery which serves as a starting point for further thought or for the solution.

Thus, of a subject who was given a mechanical puzzle to solve, it was reported: "The solution later flashed upon the subject by means of an involuntary analysis of the image" (11). Such visual imagery has been recorded for a lightning calculator who could give the sum of fifteen digits at a glance. The imagery was here apparently of an eidetic type (see pp. 361ff.). It appeared as though written in chalk on a freshly washed blackboard in the calculator's own handwriting. The calculator could, so to speak, turn to one or another imaged item. The writers describing this very special case spoke of the imagery as performing a "reference function." Thus in some

instances, the image serves as a conscious nucleus around which thought may crystallize, as when the subject analyzed the visual image of the puzzle. Here we have a center around which the thinking was organized, and marking a stage in the total process.

There is an enormous individual variation both in quantity and quality of imagery between the thinking of one person and another. Thus to the problem, "How long will it take to fill a sink 4 ft. x 2 ft. x 1 ft. 6 in. at the rate of a cubic foot a minute?" most adults will give the same answer, using much the same method; yet the imagery involved will show extreme variability from person to person.

Thought and consciousness. Many workers have stressed the fact that conscious experience is not enough to explain what goes on during thinking. It will be remembered that the Würzburg investigators maintained that the determining tendencies may work unconsciously. Many a time in these and other experimental reports we find such statements as: "The word flashed suddenly into my mind. I do not know how it got there." Such sudden flashes, without knowledge of certain intervening steps, are, as a matter of fact, common in our mental processes. It was upon just such situations that the doctrine of insight was based. Whether or not we accept this doctrine, we must accept the fact that we are unconscious of many of the processes that govern our thinking.

There exists, indeed, an intricate mechanism which is functioning during the process of thought, but which has no conscious counterpart. This fact is well brought out in an experiment where the subject was required to tie together two hanging ropes which were too short to reach each other except by special manipulation. One rope had first to be swung, and then caught on the rebound while the other was held at a slant. The experimenter sometimes gave a hint by swinging one rope, with the result that the subject then solved the problem. The subject might afterwards deny that the hint had been given. That is to say, he might be quite unconscious of the

function of the nucleus round which the solution was organized. The solution usually appeared without the subject's being able to describe the process of its formation. He was unconscious of the basic mechanism involved.

To speak of an unconscious directive or basic mechanism of thought does not of course necessitate belief in any of the more elaborate theories of the unconscious or subconscious mind. It implies only that, during thinking, ordinarily, perhaps always, unconscious processes are involved. There is no reason to believe that such processes are necessarily different in quality from those of which we are conscious. Thus, without necessarily subscribing to the Freudian theory of the unconscious mind, we can see in Freud's explanation of dreams an excellent example of the molding of thought by unconscious processes. Freud distinguishes between the manifest and the latent content of a dream. The former is the dream as it appears to the dreamer. The latter is the meaning which may be made conscious to the dreamer through the processes of psychoanalysis. The unconscious latent content regulates the conscious dream processes. Many of the Freudian interpretations are rightly felt by experimental psychologists to be overloaded by unnecessary hypotheses, but there appears to be no doubt of the soundness of this general distinction between conscious and unconscious processes in thinking.

Thus we have a picture of directive thought as proceeding from a problem, as moving by successive steps, successful and unsuccessful, to a solution, as crystallizing around nuclei of whose function we may or may not be conscious and as largely regulated by processes of which we are normally quite unconscious.

Thought and action. At the beginning of this chapter, we laid stress on the fact that thought is intimately related to action. It will be remembered that the boys on the tower were engaged in activity which persisted until the problem was solved. A series of actions which aims at a definite end is known as *directive action*. The end is usually called the *goal*.

We may use a diagram similar to the one previously used (see p. 465).

Immediate situation (→X→X→X→X→X→) Goal.

Wishing to attain a goal, the individual finds himself in a specific situation. He must then 'take steps' to attain his end. The X's will then be *actions leading to the goal.* These actions are directed towards the goal just as the trains of thought are directed towards the solution. A man is on the street when it starts to rain. His goal is to get home. He looks round for a taxi, but there is none in sight. He sees a public telephone sign, runs to the telephone and finds someone is using it. He waits till the telephone is free, and so he continues until he finally reaches home. All these actions are controlled by the fact that the goal is to be reached. The man does not perform irrelevant actions; when the storm begins he does not take a newspaper out of his pocket to read in the rain.

It is important to remember, however, that conscious processes are going on at the same time as this train of actions. In fact, we have here what may be called a train of thought acted out. Observable action is not always so prominent in our thinking, but it is almost always present, even where the simplest trains of directive thought are concerned. This is to be expected if we remember that directive thought was described as what goes on in consciousness during a process of response to a situation. Thought cannot be understood unless its intimate relation to action is continually kept in mind.

Recession of the response. It would, however, be a mistake to assume that every train of integrative thought necessarily and invariably issues in observable action. This may be seen by considering the evolution of thinking. Organic life gradually increases in complexity from the simplest forms up to human kind. As we come towards the top of the scale the period that may elapse between the situation and the response to it lengthens out, and at the same time the intermediate processes become more and more important and complex. It has been found that chimpanzees can allow a

considerable period of time to elapse under suitable circumstances between the initial presentation of the situation and the final response. With the human being this intermediate process may be lengthened still more. The observable response will at times become almost negligible, but it is generally there. The mathematician may work for weeks on a difficult problem, and finally write down the solution in half an hour. Or he may never write it down. But it is in human nature to do something about it, even in the abstract sphere of mathematical thought. Thought is ordinarily associated with observable action, though the stage of final response is sometimes omitted.

Thought and muscle. Since thought is so intimately related to response, it will not be surprising to find that during the process of thinking there are occurring widespread bodily changes of a sort which cannot be observed. In fact, one should envisage thinking and its physiological accompaniments as a series of events involving the whole organism, both physically and mentally. "We think with our muscles," it has been said. Change of muscular tension during thought is a matter of common observation. The brows are wrinkled. The posture is often intent. Even the chimpanzee may scratch his head when confronted with a difficult problem.

Investigators have lately devoted considerable attention to processes occurring in the muscles during thinking. The ultimate point at issue touches very closely the motor theory of consciousness, that is to say, the theory that consciousness, and therefore thought experience, *is* response. Into that theory we shall enter only as it involves the question of how *thought* and response are related.

In one investigation a number of subjects were trained in a highly elaborate method of physical relaxation. It is claimed that, by practice, not only can the postural muscles be relaxed to a degree previously unattained, but also the internal and external muscles of the eyes, the muscles of the larynx and the tongue involved in speech, and those connected with the

activity of breathing. There is evidence that during such complete relaxation, tension of smooth or involuntary muscles may also be diminished. When these trained subjects were in a state of complete relaxation, they reported that thought did not take place. The experience of muscular tenseness was apparently a necessary part of attention and the thought process.

In order to check these introspective reports, an apparatus was developed to detect and measure the slight electrical changes that occur in nerves and muscles during mental activity. The results are striking and fully confirmatory of the introspective reports. The records show that, during visual imagination, and during recollection and imagination of muscular actions, there was muscular activity in the eyes and the other parts of the body concerned. A subject with an amputated left hand discovered that he had no independent imagination for that hand as he had supposed, but only for the intact hand.

Many other workers have corroborated this general result. Recognizable differences have been found in the degree of tension of various muscles when subjects are preparing for tasks of different degrees of difficulty. When a subject is doing mental arithmetic, the tension of the muscles has been shown to be in a continual state of change. However, no *complete* correspondence has been found between thought processes and the *pattern* of such implicit muscular responses, as they are termed.

LANGUAGE

Language and thought. An outstanding instance of directive organic response, and one in peculiarly close relation to thought, is the activity known as speech or language. Language is behavior of a most complex and highly adaptive type. One of the chief characteristics that distinguishes it from most other action is the fact that it serves the purpose of communication. It is distinguished also by the fact that it is symbolic. It is in fact social behavior of a very highly specialized kind.

Speech, both written and spoken, exhibits many of the char-

acteristics we have attributed to directive thought. To begin with, it is plainly directive. Words and sentences do not follow at haphazard, but have a more or less direct relation to the situation—generally a social one—to which response is being made. The process of control is more particularly evident in prepared speeches and literary productions, where words and sentences are often carefully pruned, eliminated, restored, changed and so on.

Another point of similarity between speech and thought lies in the fact that in speech trial-and-error is often prominent. Most writers have had the experience of cutting out or changing a word or phrase, restoring the original word, and cutting it out or changing it back again.

Further, to say that speech is directive is to imply that it is unitary or integrative. So much insistence is laid upon unity by teachers ot oral and written composition that there is hardly need to emphasize the fact once more. The sentence, the paragraph, the story, essay or speech are all units. We know, again, from experimental researches that the word and the letter are units. This highly complex elaboration of unities within unities is effected by the integrative functioning of the organism, which is meeting a diverse situation with a unitary response.

Again, in speech, as in thought, new and old are very subtly intermixed. We have a standard system of letters to form words. These letters are common to many modern languages. The words which they form are, for the most part, peculiar to one language. Each language has a standard system of word arrangement—the rules of syntax. Much the same thing is true for spoken language, where, from a certain number of possible sounds, we fashion standard words which must be used in accordance with the grammar and idiom of the language we are speaking. Thus the raw material of language is already in our possession in a relatively fixed form before we speak or write. It is by no means new.

But the factor of newness is no less apparent. New combinations of words, newly coined words, new uses for old words, as well as such terms as intonation, rhythm, emotional and

affective factors, personality traits, all operate to make of speech a living, ceaselessly changing thing rather than a collection of stereotyped sounds. If any speaker repeated a sentence of any considerable length in *exactly* the same way, the duplication would be laid to coincidence. So important is this contrast between the new and the old in spoken communication that it has been proposed to apply the name language to the relatively static dictionary and grammar element, and to reserve the term speech for the adaptable, living drama of actual communication (2).

All these characteristics which language possesses in common with thought have been experimentally illustrated in the case of telegraphic messages. Telegraphy is of course essentially a language. Its primary elements, corresponding to phonetic or alphabetical elements, are the dot and dash, with intervals of time between them. The use and understanding of such language constitute an integrative process. A letter containing four to six elements is first perceived as a unity. Then such letters are grouped into words, which are later perceived as unities in their turn. Still later, larger language units are integrated. Further, there is a constant element in all sending and receiving of telegraphic messages, namely the identity of sounds as heard and the motions of the sender's hand. Yet there is constant intentional variation in the spacing of these sounds by the operator. The more expert the operator, the greater this momentary variation, which corresponds to inflection in speech. A telegraphic message, then, is a subtle blend of new and old. Thus language, whether spoken, written or telegraphic, exhibits remarkable similarities to directive thought.

This intimate relation between thought and language is again brought out in the studies of aphasia, particularly those of the neurologist Sir Henry Head. Aphasia is an inability to speak because of injury to the brain. Head believes that, in order that speech may take place, a further process is necessary, in addition to the fundamental process of integration. This process he calls "symbolic formulation and expression."

Formulation is the process of making a formula which is of more or less general validity; expression is the act of fitting words to it. The two activities may be distinguished by introspection, but they are intimately blended in a single activity.

A patient is told to copy the examiner's movements. Facing his patient, the examiner raises his right hand. The aphasic person of a certain type cannot repeat the movement. The examiner goes behind the patient, and both look into a mirror. The patient can now copy the movements without difficulty. The explanation, according to Head, is that in the first situation it is necessary to go through the special process of making some kind of verbal formula such as "right equals left," or "change sides." This special process is beyond the powers of the aphasic patient (4, 3). Head believes that it is the fundamental fact in the use of language. It depends on the ability to make a *suitable* formula out of the experience provided by the past and fit the proper words to it, a process in which thought and language are united in a single psychological activity.

As an outcome of his extensive experiments on remembering, Bartlett has developed the theory a stage further. Memory, he points out, tends to produce fixed patterns or schemes of thought, language and action in general. These patterns are invaluable in our everyday life: we could not eat a meal, play a game or do the simplest sum in arithmetic without them. All arithmetic is founded on the largely verbal additions and subtractions and multiplication tables which we learned at school and which form rigid modes of thought and action.

But the organism cannot allow itself to be guided by these fixed forms of thought. It must break up the patterns and recombine them for new purposes. This ability to reorganize old patterns to meet new situations is the highest psychological activity.

A key is normally used to lock a door. The schema (pattern) or habitual setting of the key thus includes the door and the activities of locking or unlocking. But it is only a robot or an imbecile who, whenever he sees a key, is immediately impelled to lock the door. A rational person may use it

for prying open the top of a can, or any one of a hundred new purposes. In so doing he tears the key out of its setting or schema and effects an adaptive integration for the purposes of the moment. So it is with words, which, stereotyped though they and their associations are, must be fused into a momentarily fresh pattern. Thus Bartlett's use of the concept of the pattern recognizes the fact of determination by the past, and at the same time avoids the charge of mechanization. In this way he apparently explains the paradox of the new solution, verbal or otherwise, which is fashioned out of old experience.

Is thought merely the action of language mechanisms? Thus language and thought are very closely related—so closely in fact that it has been held at many different times that thinking *is* language.

The general relation between thought and muscular reaction has already been considered. The particular case of the speech muscles well illustrates the general statement that no complete correspondence has been found between the patterns of mental and implicit muscular activity (*i.e.*, changes of tension or *tonus* in the muscles). By using an ingenious mechanism for analyzing tongue movements, it has been demonstrated that such movements are not universal during internal speech. When they do occur, they correspond with those of overt or audible speech for the same words in only about 5 per cent of the cases. Repetition of the same words in silent speech gives the same pattern in only 10 per cent of the cases where movement occurs.

Thus internal speech is not, as is sometimes claimed, the same thing as audible speech, produced on a smaller, inaudible scale. Even if it were, this would not mean that thought and language are identical, for thought processes can apparently be carried on in non-linguistic terms. We have all had the experience of thought for which we cannot find the words. Furthermore, animals cannot speak, yet, we shall see, it is extremely probable that they think. To equate thought and

language is not justified by experimental fact. Language is, in fact, but one form of the general response activity with which thought is so closely related.

The function of verbalization. Language does, however, occupy a unique place among our reactions. Thinking and talking have grown up together and should be considered as two aspects of the same phase of mental development. By language, thought is clarified, sharpened and made exact as by no other means. The development of modern science depends on mathematical analysis, which is a special development of language. Even on an elementary plane, a science of chemistry would be impossible without the ability to name the elements and compounds, and a science of astronomy equally helpless without the ability to name the stars. Indeed, it is evident that language, the most highly organized system of response we possess, is of the greatest advantage in our attempts to understand and dominate our environment.

The experiments described in the chapter on imagery (pp. 358f.) are instructive in this regard. It was found that the reproduction of figures is easier if they are given names of objects. In fact, when the figures were of unusual shape or relatively meaningless, they could be retained in memory *only* if they were named. It is true that the subjects differed in the name they used. Thus the same figure was called by different subjects a violin, a dumbbell, etc., as shown in Fig. 115. The point is, however, that a name was necessary if some resemblance to the figure was to be held in mind. A practical parallel to the experiments is shown when a student beginning a subject such as botany or anatomy is told what the various structures 'look like'—a procedure which is later of material help in remembering the details. One who has been told that the *cochlea* of the ear is like a snail shell, and has been shown a model, will find it difficult to forget the shape of that structure.

Other experiments have been performed from somewhat different angles. It is found that instructions of the proper kind

may improve performance in the blindfold tracing of a maze with a pencil (7). By verbalization a memory pattern of the maze may be formed by means of which many details may be recalled later. The subject may say to himself, for example, "turn to the left, once to the right, twice to the left, double the distance, then goal." Experimental investigation of the advantages conferred by the unique human capacity of speech is still, however, far from complete.

MEANING AND CONCEPTS

Mental processes and meaning. It has been mentioned that language is symbolic. This brings us to a new phase of the problem, the phase of meaning. We talk and think *about* something.

A telegrapher is taking a message. If we are to understand what is happening, we must take into account, not only the metallic sounds which a stranger hears, but their *meaning* to the operator. That is to say, for psychological understanding of what happens to the telegrapher we must put the sounds which he hears into some form of context. They cannot be understood by themselves, as mere sounds. In the operator's mind they *refer* to something.

A railway telegrapher hears the series of sounds that signal his station. He leaves the window at which he is selling a ticket, goes to the instrument, signals "MQ" (wait ten minutes) and returns. What is the *meaning* of the sounds he heard? For a subjective or introspective account we must start with the *experience* of the receiving telegrapher. It is of auditory quality and includes two sounds repeated at different intervals, making a group of perhaps eight sounds altogether. The meaning of this total set of experiences is then its context in consciousness. The verbal image, perhaps, of the name of the station calling, perhaps the fleeting visual picture of the man making the call, kinesthetic imagery of various kinds, verbal imagery perhaps of the words "must answer him," and so on.

The objective or behavioral description says that the meaning

of the group of *stimuli* is what the station agent does about it at the moment: he interrupts his conversation, walks over and gives certain movements to the telegraph key. Although these responses may vary somewhat according to the actions on which the telegrapher is engaged when he hears the signal, yet in all such different responses to the group of stimuli in question there is something similar. It is this similarity that causes us to say that the stimuli have a definite meaning.

Objectively the meaning of the light-stimulus in the conditioned reflex experiment is the response which the dog makes to the food. Introspectively the meaning of the light-experience is the idea of food. It is context in each case, but context of a different kind.

It is well known that human thinking makes very large use of signs. They enormously enhance our ability to analyze and control the environment, and at the same time very greatly economize our observable actions. These signs must be considered in relation to a meaning or context. An excellent example is found in the signs we use in mathematical reasoning. A distinction is sometimes made between the sign, where we have lost sight of the thing signified, as in mathematics, and the symbol, where we have the thing signified in mind, as in the case of words. It is largely to the highly developed use of signs and symbols that human thought owes its astonishing efficiency.

Generalization and abstraction. One feature of human thinking has captured the imagination of observers from the time of the Greek philosophers to the present. It is the fact that thought processes initiated in a concrete situation may have reference to many situations other than the original one. That is to say, a concrete thought may take on the quality of *general* validity.

A well-known experiment in physics shows that, if two weights are suspended one at each end of a uniform lever, they will balance if the products of the weights and their respective distances from the fulcrum are equal. This fact is

demonstrated for a number of particular cases, and ultimately we recognize the features that are common to them all. Somehow we have abstracted from the particular cases and see that the formula is universal. The foundations of science are built upon this ability to abstract and to generalize, to recognize common features in a variable world.

When we generalize, we are learning to respond specifically to a feature or set of features that is found in a number of different situations. This response is often, though not necessarily, a verbal one. On the conscious side, we are said to form a concept. *By the accompanying process of abstraction we learn to disregard irrelevant details.*

For example, when a child has learned to distinguish horses from other animals, it now attaches the same verbal response ("horse") to all horses, regardless of their differences. It has generalized. On the conscious side it is said to have acquired a concept, sometimes called a general idea. At the same time it has learned to disregard irrelevant detail—size, color, and so on. This is abstraction. Later, the child learns that "the three angles of a triangle are equal to two right angles." This verbal statement is made, not of any particular triangle as such, but of those features which any triangle shares with others. It is a general statement, the product of generalization. By the process of abstraction the pupil has learned to disregard irrelevancies, such as the shape or color of any particular triangle. He has developed a concept of triangle.

Many experiments have been directed upon the processes both of generalization and abstraction. We shall first consider an investigation of *abstraction.*

A group of five geometrical figures was shown to a subject for a quarter of a second, and was then replaced by a second group, and so on up to twenty-five exposures. All the groups had one figure in common (see Fig. 130). The subject was told to watch for repetition. As soon as he was sure he had seen some figure twice, he was to turn a switch which stopped the exposures. The recognition of the common figure

in a group might require any number of exposures from five to twenty-five. It was accomplished by a process of abstraction.

This process of abstraction was initiated by breaking up the group of figures. The common element was accentuated, and the others faded from consciousness. The irrelevant figures were "not merely neglected, but . . . positively cast aside and swept more or less completely from the field of consciousness" (8). Thus the process of abstraction was two-fold. There was a positive accentuation of the figure abstracted, and an equally

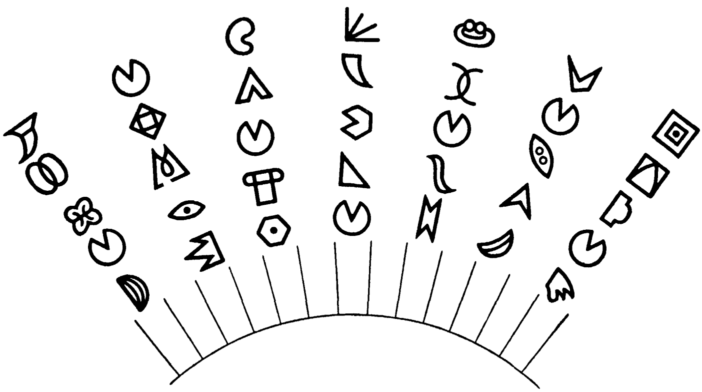


FIG. 130. DESIGNS USED FOR A STUDY OF ABSTRACTION. From MOORE (8).

positive thrusting away of irrelevants. This dual nature of abstraction points to a practical danger in everyday life. The professional specialist is obliged by the nature of his work to concentrate on certain aspects of the people with whom he is dealing. Such concentration tends to make him shut his eyes to other, equally important, aspects. The business man, who is treating people all day solely as 'customers,' often finds it difficult to remember that they are human beings. "The lawyer," said Schopenhauer, "sees mankind in all its wickedness, the doctor in all its weakness." There is no doubt that this 'professional abstraction' helps to determine the attitude of professional men towards other people.

The process of abstraction in the experiment in question was

a gradual one. The subject had first a general idea that a figure was being repeated, which gradually gave way to clearer and clearer perception of what had been repeated. This characteristic has many examples in everyday life.

In order to study the process of *generalization*, an experimenter required his subjects to repeat the names (invented) of Chinese characters. Those characters which had a part in common were given the same name. The subject gradually learned to associate the common element with the common name, and could finally give the name when the common element appeared in an entirely new context. Thus the subject learned to give a specific verbal response to a feature that was common to a number of different situations. In other words, he had generalized.

In this case, the experimenter's chief interest lay in certain practical problems. He concluded, for example, that simple settings are more efficient for the purpose of teaching a generalization. Instruction, that is to say, is best with the minimum of irrelevant detail. He also concluded that the concept can be taught apart from any setting but that the best method is to present it both in and out of the setting.

A later investigator worked with colored geometrical designs which were also named. Thus a 'vec' was the name of an equilateral triangle bisected by a line drawn from an angle to the opposite side. A 'dax' was a circle and two dots, one dot being inside, the other outside, the circle. The designs might be of any color, shape or position, provided only that the defining conditions were observed. Different varieties of the designs were exposed, and the subject was told, *e.g.*: "Each of these drawings is a dax; try to find out everything a figure must be if it is to be called a dax." He was tested by being asked to define a dax, by being asked to draw one, and by a test series in which he indicated whether or not a given figure was a dax. Similarly for the other figures. The subject learned to respond specifically to a common set of features in a number of different situations and to neglect irrelevant details. He learned to generalize and abstract. It will be noticed that

this experiment does not employ identical elements in different settings, but, instead, common features of a more complex nature.

It is important to observe that certain subjects could go through the test series without being able to give a verbal definition. The ability to acquire a concept was found to be correlated with intelligence (12).

Through all these experiments, in many respects so dissimilar, the fact emerges that the final generalization takes time to develop. Psychologically speaking, we evolve our concepts. They are not immutable. We realize that this must be so when we remember that 'generalization' and 'abstraction' are but special cases of learning. They are akin to the maze behavior of the rat which, by gradually learning to disregard the wrong passages, develops a uniform and correct response to the maze situation.

LIMITS OF THOUGHT

Thought and social environment. Thought is intimately related to social environment. The highest intellectual activities are conducted through the medium of language, which is a mode of symbolic social behavior. Even the scientist, dealing with abstract situations, must put his results into such form that others can understand and repeat his experiments.

Though his actual work may be done in solitude, without the collaboration or criticism of a colleague, the scientist's stock-in-trade is almost entirely of social origin. He starts his investigation where his predecessors have left off. In his earlier life he has acquired a general education, in school and out, in a social environment. Like that of the writer, the musician, and the pictorial artist, his thought is socially derived and socially directed. Science is a symbolic drama which the investigator plays to an audience of fellow workers.

Published scientific work probably represents the expression of the most highly logical and socialized thought. In contrast to it are many thought processes, both of adults and children

which are neither 'logical' nor couched in a form suitable for communication. It has been found that children of six and seven years of age often are unable to understand one another, a fact which suggests that the child's thinking is relatively non-logical. But it has been shown that college students, when they attempt to deal with material somewhat beyond their scope, are subject to the same limitations. Children and adults think logically and in communicable form when their knowledge and experience allow, and when their thought is not warped by prejudice or special predilection (10).

Thought in animals. The question whether animals think has been variously answered. For one who believes that thinking is internal speech, thought in animals is impossible. If one includes obscure manual (manipulative) and visceral activity in his account of thinking, he may assign thinking of a simple kind to animals, even though he does not allow them 'ideas.'

At one time this question seemed to have been settled by the theory that an animal in a new situation does not think, but struggles blindly until by chance it hits on a satisfactory response. However, the results of recent investigations are not entirely in accord with this view. Some animals solve problems by means which indicate an insight into the situation. Thus a chimpanzee was observed to sit with two bamboo canes, neither of which was long enough to reach a coveted banana. Having accidentally fitted the two canes together in the manner of a fishing rod, he jumped up and began to draw the banana towards him with the double stick (6).

The experimental work that has made possible this change of expert opinion cannot even be summarized here. It belongs primarily to the field of learning, which is here closely related to the problem of thought. It may be stated, however, that certain modern experimentalists now believe that animals show in their behavior many of the characteristics of human thought and that they think mainly through the medium of observable action, with some power of analysis and synthesis;

they talk not at all, and they act largely on the spur of the moment. Observable action in man, however, is often shrunk to a minimum through the economy of symbolic operations, and response is separated from situation by processes of prodigious complexity.

The neurologist, Herrick, says: "The human brain can fabricate symbols and abstractions; it can use language, numbers and equations, design machines, bridges, telescopes and use them. The chimpanzee does not know the meaning of $y^2 = 2px$, *and he can never find out*" (5, 290).[3]

REFERENCES

1. DUNCKER, KARL. A qualitative (experimental and theoretical) study of productive thinking (solving of comprehensible problems). *Ped. Sem.*, 1926, 33, 642-708.
2. GARDINER, A. H. *The theory of speech and language.* Oxford: Oxford University Press, 1932.
3. HEAD, H. Disorders of symbolic thinking and expression. *Brit. J. Psychol.*, 1921, 11, 179-193.
4. HEAD, H. *Aphasia and kindred disorders of speech.* Cambridge, England: Cambridge University Press, 1926. Vol. 1.
5. HERRICK, C. J. *Brains of rats and men.* Chicago: Chicago University Press, 1926.
6. KOHLER, W. *The mentality of apes.* New York: Harcourt, Brace and Co., 1925.
7. LAMBERT, J. F., and EWERT, P. H. The effect of verbal instructions upon stylus maze learning. *J. Gen. Psychol.*, 1932, 6, 377-399.
8. MOORE, T. V. The process of abstraction. *Univ. Calif Publ. in Psychol.*, 1910, 1, 73-197.
9. PETERSON, J. C. The higher mental processes in learning. *Psychol. Monog*, 1920, 28, No. 129.
10. PIAGET, J. *The language and thought of the child.* New York: Harcourt, Brace and Co., 1926.
11. RUGER, H. A. The psychology of efficiency. *Arch. of Psychol.*, 1910, 2, No. 15.
12. SMOKE, K. L. An objective study of concept formation. *Psychol. Monog.*, 1932, 42, No. 191.
13. TITCHENER, E. B. *Experimental psychology of the thought processes.* New York: Macmillan Co., 1909.

[3] Reprinted by permission of the University of Chicago Press.

PERSONALITY

A psychological study of any society shows that behavior is most usefully divided into two classes. On the one hand people show practically an identity of reaction in many situations; on the other hand the responses evoked by certain stimuli exhibit a wide range of individual differences. In our civilization almost all motorists stop at a red traffic signal and almost all Americans rise at the playing of the Star-Spangled Banner. The sociologists have referred to these common responses as customs or institutions. In contrast with these common forms of behavior we find situations in which individuals differ from one another in their reactions. On the open highway individuals vary considerably in the speed at which they drive their cars. Even in respect to observance of custom, behavior may show points of difference as well as points of identity. Almost all men in our culture shake hands when introduced to one another, but the vigor and manner of the handshake may differ.

THE STUDY OF PERSONALITY

Definition of personality. *Personality refers to the behavior of the individual which is not stereotyped either by the group or by the non-social environment. It may be defined more positively as the behavior of the individual which differentiates him from his fellows.* The term *personality* has not always been used to designate the differential behavior of individuals. Etymological studies of the word *persona* and its derivatives show a double meaning roughly equivalent to the two types of behavior just described (5). One use of the word

This chapter was written by Daniel Katz of Princeton University.

referred to the mask of the actor, when the mask indicated the role the actor was to play. Hence the one meaning of personality emphasized the assumed behavior of the individual according to his part in the drama of life. The other meaning denoted the true self or the inner man. In common modern usage in the social studies the word *person* is employed for the first of these meanings, and, therefore, in the interests of clarity the term *personality* should be used to refer only to the second meaning. Thus *person* designates the social status, or social role, of the individual; *personality* means the unique character of the individual.

The distinction between personal and institutional behavior is strikingly illustrated in an observational study of the behavior of motorists (3). The reactions of 2114 motorists were recorded at a street intersection with a boulevard stop sign. The observer noted whether (1) the driver came to a complete stop, (2) slowed down to a very low speed, (3) slowed down slightly or (4) went ahead without any alteration of speed. Fig. 131, part *A*, shows the number of motorists who exhibited these four types of behavior. A high degree of conformity is evident in the responses. About 75 per cent of the drivers came to a complete stop, 22 per cent slowed down greatly and less than 3 per cent either checked their speed slightly or failed to check their speed at all.

The behavior of motorists at a street intersection with cross-traffic but with no stop sign was then studied. The same four classifications of responses were used, but this time the cases scattered much more over the whole scale (Fig. 131, part *B*). In other words, without the institutional symbol of the stop sign, individual differences of temperament assert themselves and we no longer find a regimentation of behavior.

A question arises concerning the definition of personality as the differential behavior of the individual. Although individual differences did appear in the behavior of motorists at the street intersections with no stop signs, nevertheless the reactions observed did not differentiate the 208 people as in-

dividual personalities. People were much more alike than they were different. This definition of personality, however, does not imply a unique reaction on the part of every individual in every non-institutional situation. If some differentia-

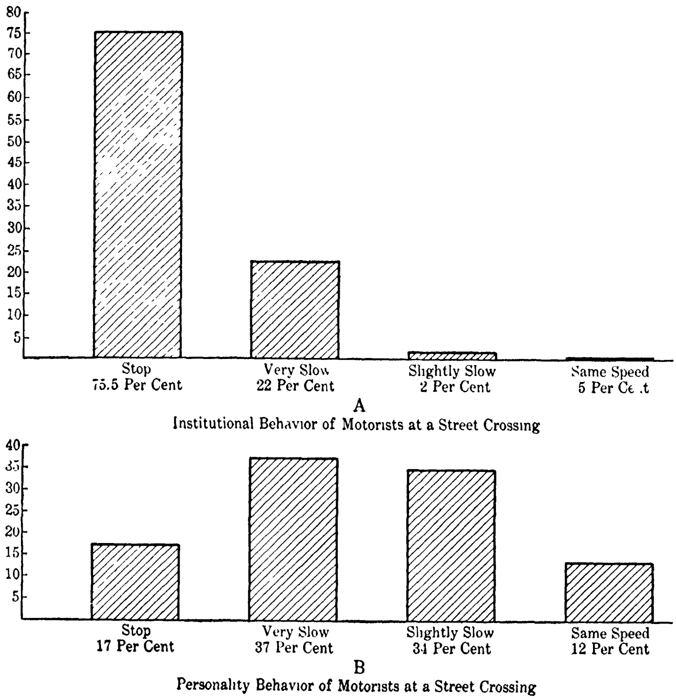


FIG. 131. BEHAVIOR OF MOTORISTS AT A STREET CROSSING.

Modified from Allport, F. H., Social and political problems. In *Psychology at work*, edited by P. S. Achilles. By permission of the McGraw-Hill Book Co.

tion occurs in a single situation, the uniqueness of an individual's personality may be found by watching his behavior in many situations. Those stimulus settings, however, which permit of practically no differentiation, can give no clue to personality. Such situations are of two kinds, social and non-social. A regimentation of behavior is produced not only by the demands of our fellow human beings but also by the non-

social environment. We all, for example, withdraw from fire and other forms of intense stimulation.

Methods in the study of personality. Personality may be described from two points of view. The one approach analyzes personality into more elementary units called traits, and these traits are studied with little attention to their interrelations. The other approach is concerned with personality as a whole. Its purpose is to study the degree of unified organization, or integration, of the individual's personality. This contrast in emphasis can be seen in everyday life. If we need to employ a trustworthy night watchman, we scan the records of applicants with special reference to the traits of reliability and honesty. If, however, we are choosing a traveling companion, we abandon our analytical search for single traits. We turn instead to the total personality pattern of the individual with whom we may have to live on fairly intimate terms for a period of weeks.

Two errors frequently occur in the study of personality, regardless of whether the emphasis is upon analysis or upon personality as a whole. The first fallacy is the popular confusion of social evaluation with naturalistic observation. We are always judging our fellows in terms of the way in which their actions affect us, rather than on the basis of their actual behavior. We refer to the courage, honesty or duplicity of acquaintances in much the same way as we call a dramatic production good or bad. Yet when we speak of a *good* play we have some realization that the excellence lies not so much in the play as in our esthetic appreciation of it. When, however, we speak of the honesty of a friend, we forget that this judgment is both a function of our own interests and experiences and of the behavior of the individual whom we are evaluating. Because we have a word or symbol for our evaluation, we tend to believe in the objective reality of the symbol by projecting it into the external world, in this case into the individual judged. We see honesty not as our own evaluation but as a quality which fairly flows from our friend.

The second great source of error is to overlook the relational implications of many alleged personality characteristics. Leadership, for example, can never be defined wholly in terms of the individual known as the leader. A leader implies individuals led, and hence leadership refers both to the personality of the leader and to the needs and demands of the individuals who are led by him. The 'great man' school of history loves to point to the emergence of such leaders as Luther, Cromwell and Napoleon and to explain history in terms of the personalities of heroes. This school forgets completely the necessity of the leader's meeting the demands of a large and powerful section of society. Men become leaders only when their programs happen to suit the needs of the times. Thus all studies of leadership which seek some constant physical and psychological characteristics in leaders miss the vital point at issue.

In spite of the confusion arising from the use of relational and evaluative concepts in the study of personality, no necessity exists for abandoning these concepts. We do need, however, to utilize our knowledge of the nature of these ideas as a precaution against error. In respect to relational notions such as leadership, it is essential to describe both terms of the relation. Instead of making leadership a function of the individual, we must seek the particular personality characteristics and the particular social situation which together make up a given instance of leadership. In respect to evaluative concepts, such as honesty, we can obtain knowledge of a fair degree of reliability by the use of quantitative methods which can be checked for internal consistency.

PERSONALITY CHARACTERISTICS

The classification of personality characteristics. Of the two points of view in the study of personality we shall consider first the *analytical*. The ideal goal of this approach is a comprehensive classification of the various aspects of personality in terms of fundamental and mutually exclusive traits.

Such a classification, however, is the last step of a finished science. Present attempts merely indicate the tendency of the human mind to oversimplify the world by an *a priori* system of nice categories and neat pigeonholes. As a matter of fact the usual classification of personality characteristics does not deal with separate and distinct dimensions of behavior, but with much the same behavior regarded from different angles. For example, the action of the football player in refusing to risk injury for the sake of an extra yard may be classified as intelligent from the point of view of the player's own interests. The same behavior, however, may be classed as cowardly from the standpoint of the glory of the player's alma mater. The characteristics of personality presented in this chapter, therefore, are described as evaluations of individuals in terms of specific purposes. Only those characteristics which have been the subject of quantitative study will be considered. Table XXI lists these evaluations of personality.

TABLE XXI

PERSONALITY CHARACTERISTICS

Characteristic	Purpose of Investigator
1. Intelligence.	Classification of individuals in relation to scholastic ability
2. Introversion-extraversion.......	Description of manner in which individuals meet their conflicts
3. Emotional stability	Prediction of mental breakdown
4. Ascendance-submission.........	Description of manner in which people deal with their fellows
5. Atypical attitudes..............	Description of people in terms of radical and reactionary ideas
6. Interests and evaluative attitudes	Description of incentives and motives

Intelligence. Intelligence is generally defined as the ability of the individual to adjust to new problems and con-

ditions of life. The essential truth of this definition is obscured by the difficult concept of adjustment. Adjustment really involves both intellectual and emotional factors. Moreover, the problems in life to which we try to adjust generally arise out of our social environment. Now, the social environment consists largely in our reactions to one another. Hence it is constantly shifting, and our very act of adapting or failing to adjust changes the social environment. The genius who refuses to adjust to the ideas of his time may by this act start social changes which result in the adoption of his so-called maladjusted attitudes years after his death.

In view of the host of problems connected with the concept of adjustment, a definition of intelligence might well confine itself to the psychological sphere and omit this sociological criterion. Psychologically, intelligence is a term for the ability of the individual to learn and think. In the thought process implicit reactions, such as subvocal speech and incipient eye movements, are frequently substituted for overt bodily responses. Implicit reactions are thus symbols for gross bodily activity and enable the individual to try various solutions of a problem without recourse to overt trial-and-error. Intelligence is the ability of the individual to solve problems in terms of implicit reactions. The intelligence tests measure the ability to use one type of substitute response in the solution of problems, namely the language reactions of the schoolroom.

A well-known intelligence test, the Army Alpha scale, is illustrated in Table XXII. The Army Alpha is a battery of eight tests given to United States army recruits during the World War. For a clear understanding of it, we must examine the original basis upon which the first tests of intelligence were constructed. The first widely used scale for measuring mental ability was the work of Binet, who after much trial-and-error experimentation produced a test in 1905 which became the groundwork for all later scales. Binet's success consisted in the application of two important principles: (1) the use of a large number of tests which conceivably might differentiate individuals according to gradations of ability, and (2) the discovery

of an objective criterion against which the differentiation of his scale could be checked. The criterion was the differences in ages of a group of children. Obviously as children grow older they develop mentally. The problem, then, was merely to correlate the age differences with the differences in performance in scholastic problems. Problems that did not differentiate children according to their ages were discarded.

Binet's scale has been revised and standardized for general use. Standardization is the process of finding the significance of the items of a test in terms of their solubility by large numbers of subjects. In the revision of the Binet scale, norms were found indicative of the average performances of children of various ages. Reference to these norms, or average performances, gives the mental age of a child. Thus a child has a mental age of ten, if he equals the performance of the average ten-year-old. The intelligence quotient or I.Q. expresses the relation between the child's mental age and his chronological age. It is found by dividing the mental age by the chronological age and multiplying the quotient by 100 so that it is stated in whole numbers. If a child of ten has a mental age of twelve, his I.Q. would be 120.

A knowledge of the construction of the intelligence test enables us to evaluate the following widely accepted generalizations of certain test results: (1) the growth of intelligence stops somewhere between thirteen and one-half years and sixteen years, (2) the intelligence of the average adult in the United States is equivalent to that of the child of thirteen and a half years and (3) intelligence is inherited.

The first proposition that mental growth stops somewhere between thirteen and a half and sixteen years is really a statement of the limitations of the tests. Since the tests originally came from items which differentiate school children, they do not show any progressive differentiation of adults of various ages. No real evidence as to the time of intellectual maturation is offered. Since most people leave school between the ages of twelve and sixteen, their subse-

TABLE XXII

The Army Alpha Scale

Sample Items

Test 1. Directions.

Item 12. If six is more than four, then, cross out the five, unless five is more than seven, in which case draw a line under number six.

1 2 3 4 5 6 7 8 9

Test 2. Arithmetic.

Item 10. If it takes 8 men 2 days to dig a 160-foot drain, how many men are needed to dig it in half a day?

Test 3. Common Sense.

Item 9. Why are warships painted gray? Because gray paint
___ is cheaper than other colors
___ is more durable than other colors
___ makes the ships harder to see

Test 4. Antonyms and Synonyms.

If the two words of a pair mean the same or nearly the same, draw a line under *same*. If they mean the opposite or nearly the opposite, draw a line under *opposite*.

Item 11.	commend	approve	same——opposite
Item 19.	adversary	colleague	same——opposite

Test 5. Sentence Rearrangement.

Think what each of these sentences would say if it were straightened out, but don't write it out. If what it would say is true, draw a line under the word *true*; if what it would say is false, draw a line under the word *false*.

Item 4.	east the in rises sun the	true——false ✓
Item 22.	happiness lists great casualty cause	true——false

TABLE XXII *continued*

Test 6. Number Relation.

Complete the number series.

9	1	7	1	5	1	3	1
4	5	8	9	12	13	16	17
1	4	9	16	25	36	38	51

Test 7. Analogies.

Underline the correct word.

Item 15. cold—ice :: heat— wet cold steam stars
Item 18. lead—bullet :: gold— paper coin silver copper

Test 8. Information.

Underline the correct word.

Item 5. Turquoise is usually yellow red green blue
Item 16. Rodin is famous as a poet painter sculptor composer

quent mental development expresses itself in earning a living, in marriage and in other relations. To say that their mental development stops, because they do not grow increasingly adept at the arithmetic and language problems of grammar school, is to forget that the four walls of the schoolroom do not include all that there is of life.

The second conclusion, that the mentality of the average individual in the United States is that of a child of thirteen and a half years, is on its face logically absurd. What it means is that the army recruits in the World War did, on the average, no better in the test which measured early scholastic aptitude than the thirteen- or fourteen-year-old child. The same objections raised to the assumption of early mental maturity hold here. Tests built to differentiate school children are not going to differentiate adults satisfactorily, especially since the adults have spent their time since leaving school on other problems than those of the school bench.

The third conclusion, that the intelligence tests measure innate mental ability, is based in good part upon the constancy of the I.Q. Tests made of the same children at various ages show that there is little variation in the I.Q. Children who are bright at six are still bright at eight, and children dull at six are still dull at eight. Since inheritance is constant while the environment may vary, it is inferred that the constancy of the I.Q. can be accounted for only by inheritance. Two possibilities have not been taken into account, however.

The first is that prenatal development as well as the early growth of the infant are alike the result of both inheritance and environment. The fact that given structures (possibly responsible for the I.Q.) are fairly well fixed at an early point in life does not mean that these structures have no developmental history. The second possibility is that the environment is more constant a factor than has been realized. Children whose parents are well educated have an advantage in acquiring verbal techniques over children whose parents are uneducated. And for the great majority of children this significant feature of the environment remains constant.

That this second consideration concerning the importance of the environment is more than a possibility is suggested by a number of experimental studies. For instance, a study has been made of the effects of the home environment on adopted children (8). A group of seventy-four foster children were tested before and after residence in foster homes. These children were given tests when they were first adopted, and after four years in their foster homes they were retested. The average intelligence quotient for the group increased significantly as a result of the more favorable environment of the foster home. Moreover, the children placed in superior foster homes showed a greater gain in I.Q. than those placed in poorer homes. In the same investigation, a group of one hundred and twenty-five pairs of siblings, *i.e.*, children of the same parents, was tested. One member of each pair had been placed in a superior home and the other in an inferior home. The I.Q. of the former

averaged nine points higher than the latter. The children had been separated about seven years on the average.

Another line of evidence for the effect of environmental factors upon the I.Q. comes from a study of identical twins. If intelligence tests measure only inheritance, then identical twins should test alike regardless of the conditions under which they are reared. Identical twins, however, when reared apart, average twice as great a difference in I.Q. as identical twins reared together. The average difference in I.Q. for four pairs of identical twins, raised apart, was 10.9 points; while fifty pairs of identical twins, not separated, averaged a difference of only 5.3 points (18).

Intelligence tests, therefore, measure both inherited and acquired ability. We have no method of knowing to what extent each factor enters into test performance. This fact invalidates the elaborate efforts to establish innate differences between races and between classes in any one society by means of intelligence tests.

Although we cannot accept the early conclusions of mental testing regarding the growth and inheritance of mental ability, nevertheless this field offers a number of contributions. On the side of method, intelligence testing has worked out statistical techniques of value in the study of personality problems.

One finding of significance for psychological theory contributed by the tests concerns the nature of mental processes. The statistical results indicate that scholastic aptitude is not a composite of many independent specific skills. The ability to deal with language materials is fairly well generalized. In an intelligence scale, all the individual subtests are positively correlated. For example, subjects who do well in the arithmetic problems of the Army Alpha scale generally do well in the analogies test, the test of disarranged sentences and the other subtests of the scale. In fact it is difficult to find a set of problems, involving the manipulation of language symbols, which are negatively correlated. Obviously, positive correlation coefficients would not be universally found between different types of mental measurement, if the tests were measuring un-

related and specific abilities. This does not necessarily mean that scholastic aptitude is a single unitary capacity. Although it involves one dominant common factor, it is probable that other general and specific factors are also represented by the final test score, or by the I.Q.

From the practical standpoint the intelligence tests have justified their use in industry and in education. In industry an employer is concerned with keeping down labor turnover by selecting the right man for the right job. Vocational tests built along lines similar to the intelligence scales have proved an economic saving, not only by eliminating inferior workers from positions for which they are not equipped, but also by keeping superior workers out of inferior positions at which they would not stay.

In education, intelligence tests are a convenient classificatory device for grouping students according to ability. Homogeneous grouping simplifies the problems of teaching and makes possible special types of instruction for various groups. Under the older system of non-segregation, the feebleminded child, discouraged by his consistent inferior performance to that of younger children, often became a truant and a delinquent. Special classes for the feebleminded in the manual arts have saved many of them from delinquency. Another use of the mental test has been its diagnostic aid with problem children. Behavior problems and school failures result from emotional causes and from superior as well as inferior scholastic ability. The intelligence test may help to discover whether the maladjustment lies in the intellectual sphere. A third use of the mental test in education is to furnish an added criterion for admission to secondary school or college.

Introversion-extraversion. A growing field of research concerns itself with personal differences in emotional adjustment and emotional orientation. The emotional adjustment of the individual is closely tied up with his manner of meeting conflicts. Hence a fundamental interest in personality research centers about the problem of how individuals attempt to solve

those situations in life which impel them to opposed courses of action. This interest has led to a description of personalities as extraverted or introverted.

The introverted personality meets life's problems by turning from active participation in the objective world to an inner world of thought and phantasy. The extraverted personality turns from an introspective consideration of his problems to overt action. This classification is thus definitely related to the popular discrimination between the practical man of action and the idealistic visionary. Most psychologists, however, do not regard the introverted personality and the extraverted personality as sharply contrasted types into which all human beings can be classed. Rather they believe that almost all individuals possess both introvertive and extravertive characteristics, so that an introvert would be one who has more introvertive than extravertive habits.

A review of the studies in this field shows that three main aspects of introversion-extraversion have been emphasized by various writers (10). One emphasis is upon the direction of interest of the individual, that is to say, whether the individual is self-centered or interested in the world around him. The second conception refers to ease of social adjustment. The extravert adjusts to social situations more readily than the introvert. The third notion is concerned with the emotionality of the individual. The introvert expresses his emotions less freely than the extravert, and this inhibition or blocking makes the introvert more sensitive emotionally to a wide range of stimuli. These three points of view are not mutually exclusive. Rather they supplement one another to make up the final generalization known as introversion-extraversion by describing its intellectual, social and emotional phases.

The psychological mechanisms in introversion-extraversion. The degree to which an individual is introverted or extraverted is a matter of his specific developmental history. E. B. Holt has explained introversion-extraversion genetically on the basis of his description of the development of behavior

in the individual. In this process two basic reactions can be observed: *adient* responses and *avoidant* responses.

Adient responses give the organism more of the very stimulation which called them forth. Outreaching, inquiring, examining and grasping responses are examples of adient reactions (12). The basis of adience is the reflex circle which is described on page 428. *Avoidant* responses, on the other hand, take the organism out of the range of the stimuli which originally evoked them. They are produced by intense or over-strong forms of stimulation.

These two fundamental reactions are the clue to extraversion-introversion. The introverted character is the sensitive organism that has developed many avoiding reactions in comparison to the extravert. Extreme introverts shrink or withdraw from stimulation to which the normal person has only approaching or adient responses. The extreme extravert (Holt calls this type the aggressive character) is characterized by many outgoing or adient responses. He meets obstacles by a direct frontal attack. When confronted by problems that cannot be demolished by sheer aggression, he is incapable of solving them. On the other hand, the introvert falls back upon imaginal processes when he meets problems. Implicit responses or ideas take the place of overt activity. The introvert, however, does not compare his ideas sufficiently with the realities they symbolize to make him an objective thinker. He is more likely to become a poet or artist. And if his introversion is extreme, he may lose complete contact with reality and think only in terms of the phantasy of the mentally disordered patient.

Whether a stimulus is sufficiently intense to produce an avoidance response depends upon the physiological make-up and experience of the organism. A given form of stimulation which is over-strong for a child of six months may be mild for a child of two years. Introverted habits of withdrawal and avoidance can develop in the child through contact with a harsh and rigorous environment with which it is unable to cope. From the blows and scorn of older and stronger children the child may withdraw within himself and construct a

dream world of imaginative play. Nutritional and glandular factors which contribute to the structural development of the organism also help to determine the degree of sensitivity of the child to various types of stimulation.

This developmental view of introversion-extraversion helps us to understand the different forms which introversion assumes in different individuals. If the intense stimulation which produces avoidance in the child is predominantly social, as for example continual parental admonition, the child's subsequent shrinking from stimulation may be largely a withdrawal from social situations. As an adult he may find difficulty in adjusting himself socially but he is not necessarily introverted in respect to the non-social environment. Inventors and scientists often represent this type of introversion. Since, however, the non-social and the social environments are so closely related, the more usual occurrence is an avoidance of all reality whether social or non-social.

The results of tests of introversion-extraversion confirm the description of this trait as an avoidant or adient orientation toward one's environment. A number of investigators have found that physical handicaps are associated with introversion more frequently than with extraversion. Such handicaps as partial deafness and crippled limbs limit the child's exploration and manipulation of his environment and encourage the construction of an inner imaginal world.

Tests of introversion-extraversion also show that women tend to make slightly higher introverted scores than do men. This finding does not establish any innate differences between the sexes but it reflects the different training of men and women in our civilization. From an early age girls are hedged about with more restrictions than boys. They are coddled more than their brothers and at the same time fettered with more taboos. These early inhibitions make women less objective and more personalized than men.

A related finding is that introverted students are more interested in journalism, literary pursuits and medicine, whereas

extraverted students are more interested in engineering, law and architecture. One study indicated a greater introverted trend among literature majors in college and a greater extraverted trend among science majors. Occupational groups, when tested, also show slight differences in introversion-extraversion scores. Thus executives, foremen, policemen and salesgirls were in general on the extraverted side of the scale, whereas clerical workers, accountants, research engineers and teachers were on the introverted side.

Emotional stability. The characterization of personality in terms of introversion-extraversion has for its purpose the description of the manner in which individuals meet their conflicts. A more practical emphasis upon the emotional phases of personality is concerned with the possibilities of mental breakdown in supposedly normal individuals. For this purpose, characteristics of both extreme introversion and extreme extraversion are considered together and personality is described from the point of view of *emotional stability*. The interest here is not in the type of mental disorder but in the extent of the neurotic tendencies of the individual which make for breakdown.

During the World War many soldiers suffered from shell shock and other forms of nervous disorder. There were large individual differences in the ability of the men to stand up under nerve-wracking experiences. The United States government became interested in a method for selecting emotionally stable recruits for duty abroad. The Woodworth Personal Data inventory was the result. This test lists in question form two hundred neurotic symptoms found in the case histories of mentally disordered patients. The questions are to be answered by *yes* or *no*, and the number of symptoms, thus discovered, determines the subject's neurotic score.

Recent research has improved the diagnostic value of the psycho-neurotic inventory, or emotional stability test. Mental hygiene clinics and personnel offices have found many maladjusted individuals through its use. In one study of the validity

of this test, clinical records of 250 subjects were compared with their scores on a psycho-neurotic inventory (11). Serious cases of maladjustment on the basis of clinical observation showed scores in the test far above those made by the average individual. The neurotic inventory thus serves a useful function in selecting from large groups individuals in need of psychiatric attention, but it does not select out all such cases.

Ascendance-submission. The popular division of mankind into leaders and followers is paralleled in social psychology by a description of individuals as ascendant and submissive. Ascendance refers to the domination and control by one individual of his fellows in face-to-face situations (1). It differs from the concept of leadership in that leadership includes control by means of the cloak of authority and by superior social status as well as non-institutional domination. Ascendance-submission is another way of describing much the same behavior covered by the notion of introversion-extraversion. But whereas the purpose of introversion-extraversion is to characterize individuals by their manner of meeting their problems, ascendance-submission emphasizes the way in which individuals deal with their fellows.

The behavior of animals, other than man, can also be described in terms of ascendance-submission. In one experiment a limited food supply was presented to fifteen pairs of previously unacquainted monkeys (16). The percentage of the food obtained by each animal over a period of thirty such presentations was taken as a measure of dominance. Other forms of social behavior were found to go together with this measure of food behavior. Regardless of gender, the dominant animal played the masculine role in sex activity, initiated fighting and play and was more active. The submissive animal played the feminine role in sex activity and responded to the aggressive behavior of the ascendant animal by passivity, cringing or flight.

One test for ascendance-submission pictures a number of specific situations from everyday life. The subject is asked

to record his customary behavior in these situations. Two items from the test follow (4):

1. At church, a lecture, or an entertainment, if you arrive after the program has commenced and find that there are people standing but also that there are front seats available which might be secured without 'piggishness' or discourtesy, but with considerable conspicuousness, do you take the seats?

habitually .
occasionally .
never

2. When you see someone in a public place or crowd whom you think you have met or known, do you inquire of him whether you have met before?

sometimes
rarely
never

This test has one advantage over many other tests of personality in that it is a verbal representation of specific and real behavior situations. People are more accurate in their report of their actions in specific situations than they are in their answers to general questions such as, "Do you express your emotions freely?" The specific question, moreover, is much more likely to convey the same meaning to all subjects than is the general question.

The scores made by subjects on the ascendance-submission test and on tests of introversion-extraversion show a positive relation between extraversion and ascendance. Correlation coefficients ranging from 0.15 to 0.51 have been found between these traits. Although these coefficients are not large in themselves, they are all positive and they are large in relation to the reliability of these tests.

In one study of face-to-face leadership (ascendance) leaders and followers were selected from three groups. Criminal followers, privates in the United States army and student followers were compared with criminal leaders, non-commissioned army officers and student leaders. The following traits

were common to all three groups of leaders: self-confidence, speed of decision and finality of judgment.

Is ascendance a consistent trait of an individual, or is it a function of the situation and the individual's attitude at the moment? Undoubtedly the situation and momentary attitude are important in determining ascendant and submissive behavior. Most of us are not aggressive until a genuine interest or a streak of personal vanity is touched. Then we become highly ascendant. Nevertheless, if we compare the behavior of many individuals in many situations, we find a characteristic aggressiveness or submission displayed by the same individual more frequently than we should expect from chance. Controlled observational studies of preschool children show definite types of ascendance even at this early stage of development. Certain children consistently control their playmates through brute force, and others show a consistent leadership through indirect and artful suggestion.

Personality attitudes. Although no sharp line of demarcation can be drawn between attitudes and personality traits, a distinction can be made for practical purposes. Personality traits refer to characteristic forms of behavior; attitudes are mental sets for certain kinds of verbal response expressing value (14). The avoiding behavior of the introvert is an expression of a personality trait, whereas the verbal affirmation of socialistic ideas is indicative of a radical attitude. Attitudes refer more to ideas, traits to overt behavior. The subject-matter of social psychology in large part is concerned with attitudes. In this field are included opinions on social and political issues, religious beliefs, racial and national prejudices, occupational and professional interests and attitudes toward moral standards. Here social psychology joins hands with the social sciences in attacking the same problems of social interaction.

The approach of the social psychologist differs, however, from that of the social scientist, in that the former is concerned with the mechanisms of social interaction, whereas the latter

is interested in the content carried by these mechanisms. Thus the social scientist does not study the nature of attitudes as such, but the particular attitudes expressed by an individual in a certain situation. He is concerned with the meaning of the political candidate's speech in respect to his stand on the tariff, on taxation, etc. The social psychologist, on the other hand, is interested in the speech as verbal symbols which call out conditioned emotional reactions in the audience. In a propaganda campaign, the student of the social sciences would know *what* appeals might be used in different sections of the country; the student of social psychology would know *how* these appeals should be manipulated to produce the desired results.

A comprehensive study of attitudes would take us far beyond the field of personality, because so much of our social ideology is stereotyped or institutionalized. For example, our judgments of our fellows on the basis of their racial origins, their economic status and their religious affiliations are regimented by our social environment. Such racial, economic and religious evaluations are much the same for a given culture area and show few differences of personality (14). In fact in the social sciences they are technically known as *stereotypes* or fixed pre-judgments. Where social influences are less unified in their effects upon the individual, personality attitudes have an opportunity to develop.

It often happens, moreover, that individuals develop conflicting personality and institutional attitudes in respect to the same questions. A careful study of an American village disclosed that the townspeople had two sets of attitudes and practices in respect to many church and community matters (19). As good church members they thought they must regard baptism as real in the sense that the relationship between God and the baptized was altered. As private individuals, they regarded baptism as symbolic.

The subject about which an individual expresses an attitude, therefore, is no absolute guarantee that his attitude is either institutional or personal. Two types of attitudes that generally

show personal differences, however, are atypical opinion and interests.

Atypical opinion. Radical and reactionary attitudes are deviations from the accepted beliefs of the time. A radical idea implies a change in the direction of the new; a reactionary view emphasizes a return to an older day. No idea is radical or reactionary save in relation to a particular culture at a particular time. The term *atypical* is used to include both radical and reactionary ideas. Research findings have confirmed the common view that those who hold atypical opinions are generally more emotionally convinced of the correctness of their views than those who support less extreme doctrines (2).

In common thought, radical and reactionary ideas are linked to distinctive personality traits and personality types. The radical, for example, is regarded as the individual who rationalizes his own inferiority in terms of the present evil system of things (1). Research findings, however, do not confirm this point of view. The bulk of evidence shows no characteristic differences between radicals and reactionaries in intelligence or in emotional traits. Rather it suggests that the significant differences lie in cultural origins.

Although individuals vary in their degree of atypicality from one issue to another, it is not difficult to find individuals who maintain a consistently radical or reactionary attitude in different situations. Such consistency appeared in one study which covered the following social, political and economic issues: birth control, miscegenation, the powers of labor organizations, socialization of medicine, academic freedom, divorce, the protective tariff, our Latin-American policy and social limitation in mating (23). Extreme radicalism on one question was more often than not accompanied by extreme radicalism on the other issues. Quantitative study has also confirmed the popular impression that people who accept a radical economic philosophy are generally radical in their religious views.

One reason why radicalism cannot be related to distinctive personality traits is that individuals embrace the same radical

doctrines for widely differing reasons. It is useful, therefore, to distinguish between the people who are in revolt against institutionalism in general and those who are opposed only to our dominant social institutions. One man is in the radical movement because he is opposed to all forms of authority; his colleague may be there because of his specific objections to present social forms and his definite program for replacing them. The difference in radical motivation helps to explain the observation of Walter Lippmann that the labor movement preaches solidarity but seems to propagate by fission. The chronic rebel is continually breaking away from every organized form of revolt as soon as it reaches a fair measure of organization.

Interests and evaluative attitudes. Achievement is a function of many factors, but psychologically the two most important are ability and motivation. Individuals differ from one another in the type of activity which interests or motivates them. Laboratory science, involving the manipulation of apparatus and materials, may spur one student to energetic performance, but may leave his philosophically minded companion bored and lethargic. Probably more vocational maladjustment is due to lack of interest in a particular kind of work than to lack of ability, although interest and ability in a given field generally go hand in hand. Two questions arise concerning the nature of interests. How stable, or how permanent, is an individual's preference for a certain activity? And how specific is his interest to a single situation?

Experimentation has demonstrated the stability of many of our interests. Secondary-school and college students have shown considerable permanence in their likes and dislikes over a period of years. In one study it was found that students' vocational choices remained the same in three-fourths of the cases on a retest two years after the original test. Another investigation showed that items of occupational interest successfully differentiated engineers, lawyers, salesmen, ministers, physicians and school teachers at 25, 35, 45 and 55 years of age (21).

A conclusion from the same study indicated that, in general, the things people like most at 25 years of age are liked better and better with increasing age, and the things liked least at 25 years are liked less and less.

In the studies of stability of preferences, interest has been measured by the subject's recording *like, dislike* or *indifference* for several hundred items which cover occupational characteristics, types of amusement, peculiarities of people, school subjects and miscellaneous activities and objects.

Are interests specific to isolated items, or are they generalized to cover wide fields of activity? Everyday observation and controlled tests agree that both specific and general interests exist. Thus, individuals who have a general interest in mechanical things have been differentiated by tests from individuals who have a general social interest (9).

An interesting attempt to portray personality in terms of general interests is Spranger's classification of evaluative attitudes (20). Spranger describes six basic types of interests which he believes color almost all our mental activities. They are (1) the theoretical, (2) the economic, (3) the esthetic, (4) the social, (5) the political and (6) the religious.

The theoretical or intellectual interest is the preoccupation with observation, reason and the discovery of truth. Plato exemplified this attitude when he banished poetry from his ideal philosophic state. And the mathematician revealed the theoretical attitude in his question upon hearing a Beethoven symphony, "Beautiful, but what does it prove?"

The economic attitude is the emphasis upon utility as against all other values. Knowledge is evaluated in terms of its direct applications to life's problems. People are judged on the basis of their earning capacity. The proverb "Honesty is the best policy" is an example of the practical man's attitude toward ethics. The miser represents the economic interest run wild.

The esthetic attitude places its value upon the life of the imagination. The interest here is upon form, beauty, harmony and proportion. The esthetic aim is one of self-realization and

self-fulfilment. The idea of use in regard to an esthetic object, whether technical or moral, is foreign to such an attitude.

The social attitude in its highest development is the love of one's fellow men. We find the social attitude in the gregarious tendency to foregather with others for no purpose beyond the pleasure of association. The extreme social individual lives vicariously in the experiences of others. The communistic communities of the Utopians represent the social ideal. Whereas the economic attitude stresses self-preservation, and the esthetic, self-realization, the social interest emphasizes self-sacrifice.

The manipulation of his fellow men is the chief interest of the politician. The political value-attitude is not confined to the politician but can be found in all leaders who desire ascendancy over others. The successful leader must be a realistic student of human nature, for he "must take people as they really are, while the pedagogue is inclined to see them as they might be" (20).

To fathom the final secret, to understand the ultimate meaning of life, characterizes the religious attitude. Such understanding goes beyond knowledge and rests upon belief and faith. Science deals with the finite and measurable, religion with the infinite.

These six attitudes, Spranger believes, are found in varying degrees in all personalities. To understand people it is necessary to know which interests are dominant in their make-up. A test entitled the *Study of Values* has been devised to determine the relative prominence of these basic interests (22). In Part I of the test the subject is asked to check one of two alternative answers to statements designed to force a choice of interests. Question 1 from Part I of the test follows:

> The main object of scientific research should be the discovery of pure truth rather than its practical applications.
>
> (*a*) Yes; (*b*) No.

In Part II, one of the four alternative answers is to be chosen to indicate the subject's attitude toward the situation described. For example, Question 14 reads:

If you should marry, do you prefer a wife who
... (*a*) can achieve social prestige, commanding admiration from others;
... (*b*) likes to stay at home and keep house;
.... (*c*) is fundamentally spiritual in her attitude toward life;
(*d*) is gifted along artistic lines.

The whole test contains 120 possible answers, 20 for each of the six values.

The *Study of Values* is a valid means for determining group differences in generalized motives (7). Although its validity for individual prediction is not perfect, it has successfully differentiated groups of subjects who have varying occupational interests. This test has also been compared with the following experimental test of interest. Twenty-two newspaper items were selected from a large number of newspaper clippings by four judges as the clearest representations of Spranger's six types of value. A group of naive subjects glanced over these items and later were given a memory or recognition test for the items. At a subsequent date the subjects were given the *Study of Values* test. The results showed a fairly high correlation (0.71) between the group scores on the *Study of Values* and the recognition of types of news items.

The measurement of personality characteristics. In discussing the traits of personality, reference has been made to tests of introversion, ascendance, emotional stability and value-attitudes and to the results of these tests. The construction and application of the intelligence test have already been described, but how do psychologists measure the non-intellectual traits? What are the methods for studying people which give more accurate information than our everyday casual observation?

Almost all techniques of personality testing reduce to two basic methods. In the first method, the actual behavior representative of a given trait is measured directly as it occurs in social situations. The problem here is to obtain some measure

of control over the situation without destroying its validity as a natural event. Control is necessary so that we can tell whether the situation remains the same for all individuals studied. One form of control consists of complete observational records with the use of concealed dictaphones and of movie cameras operating through one-way screens. Another form of control lies in staging a situation in such a way that the subject is unaware of the staging. In one experiment, for example, appeals were made to students for contributions to various causes. The appeals were staged as if they represented a genuine drive for funds.

When this first method is used, it is necessary to sample enough of the behavior in question to obtain an adequate picture of the trait tested. In other words, we need to know how many test situations to take in order to obtain an accurate estimate of the behavior in which we are interested. This can be found by the *internal consistency* of a battery (series) of tests, that is to say, the extent to which the subtests in the battery agree as measured by their average intercorrelation. For this purpose coefficients of correlation between all components of the battery are calculated. By taking the average of these correlations, we have a measure of the internal consistency of the battery. If this average intercorrelation is low, say under 0.40, it may mean that we have not sampled enough situations in which the behavior in question is manifest. By adding more subtests we can probably increase the internal consistency of the test. If additional situations do not raise the average intercorrelation, then the behavior tested does not represent a unitary personality trait.

The second method of personality testing is to measure responses which supposedly are valid substitutes for the actual behavior in which we are interested. This method has been used for measuring introversion-extraversion. Instead of observing actual introverted and extraverted behavior, paper and pencil situations have been set up which are assumed to be indirect indices of introversion-extraversion. Subjects are asked to answer a standard list of questions, as:

Do you like to sell things?
Do you day-dream frequently?
Are you inclined to keep in the background on social occasions?
Do you express such emotions as delight, sorrow, anger, etc., readily?

The answers are tabulated and scored for large groups of subjects, so that norms or standards exist by which an individual's position can be stated in reference to certain groups of the population.

When this second, or indirect, method of personality testing is employed, the tests must be *validated* to insure a genuine measure of the personality trait in question. The *validity* of a test refers to the closeness of its correlation with some criterion admittedly representative of the personality trait. Thus a test is valid to the extent that it really measures the variable it is supposed to measure. Intelligence tests, for example, are valid tests, not of general intelligence, but of general scholastic ability, because differences in scholastic achievement are the criterion against which they have been validated.

Various criteria have been used against which to validate personality tests. A common criterion is the rating or judgment of individuals by their associates. A group of acquaintances are instructed to rate one another on the degree of the trait possessed. The average estimation or rating of the subjects can then be compared with their standing in test scores. By experimenting with the items of the test, it is possible to obtain a final test which will place the individuals in the same relative position in the group as their average ratings place them. This criterion of judgments by associates may suffer from a number of errors. One error is the central tendency of judgment in which individuals tend to underestimate large quantities and overestimate small quantities. Thus we tend to underestimate the brilliant and overestimate the stupid. Another error is the 'halo' tendency, in which a strong impression made by one trait of an individual colors the judgment of his other characteristics. The teacher is so influenced by the

obedient deportment of a pupil that she overestimates his intelligence, his honesty and his other traits.

Careful selection and training of raters can reduce these errors to negligible proportions. At its best, validation by pooled ratings of judges means that a test score of an individual classifies him according to expert social opinion. When we are dealing with personality characteristics, the basis of which is largely social evaluation, competent ratings furnish the most desired form of validation. With personality traits, which have a more objective basis, more objective criteria can be found for validating a testing program.

THE ORGANIZATION OF PERSONALITY

General vs. specific traits. The description of separate personality traits and attitudes is basically an analytical approach to the field of personality. It is also possible to study personality from the point of view of the individual's total pattern of behavior. In this approach two questions arise. The first question is really a logical extension of the problem met in the discussion of intelligence and interests. Are the separate traits of personality really distinct entities, or are they expressions of a unified pattern which colors everything the individual does? How thoroughly integrated is the personality? The second question asks: if any relationship does exist among personality traits, how is it to be explained? Here the emphasis is not upon the extent of integration but upon the way in which integration develops.

The traditional concepts concerning personality integration and personality types present opposed views. The specificists hold that behavior depends upon the particular situation in which the individual is placed, and the specific training of the individual to the various elements in that situation. They assert, for example, that the student who walks off with a library book is not necessarily untrustworthy with respect to money. On the other hand, the supporters of the idea of generality assume that responses to varying situations are deter-

mined by a general function of personality, fairly independently of the specific nature of the situation. Unfortunately, these traditional concepts have obscured the amount of genuine agreement in the experimental findings on the problem.

The factual evidence concerning the extent of integration of behavior is in essential agreement. Individuals show a high degree of consistency in their behavior, but not a complete consistency. It will be remembered that in the experiments described on pages 460f. concerning the consistency of expressive movements no general factor of motility was found but certain functions were very closely related (6). It will also be recalled (pp. 507f.) that the subtests in a scale for measuring scholastic aptitude show too high correlations to be measurements of independent specific abilities.

The *Study of Values*, the test for measuring Spranger's generalized interests, has given results indicative of the consistency of attitudes (22). In this test the subject is presented with a number of problems and situations to which he can respond either practically, theoretically, esthetically, religiously, socially or politically. The specific content of the situation and problems varies, but the general evaluative approaches are kept constant. Subjects who show a practical outlook on life in one situation tend to maintain the same attitude in other situations. A similar consistency is found in the other value attitudes.

From experimental findings, therefore, personality should not be regarded as a composite of many unrelated, wholly specific characteristics. On the other hand, individuals are too complex, they have too many water-tight compartments in their thinking and acting, to validate the theory of a personality type as a general unitary factor pervading the whole personality. An individual can be classified as possessing not one but a number of generalized functions. General types of behavior exist, but no one general factor accounts for all the conduct of an individual. We can describe an individual in terms of such dimensions as mechanical ability, scholastic ability, introversion, etc., but not in terms of a unitary type of personality.

The nature of the consistency of observed behavior is a

matter of psychological interpretation. The real psychological issue at stake is whether such consistency arises through a learning process in which elements are pieced together to make up the total pattern, or whether general functions exist from the start and are incapable of analysis. Many psychologists believe that to attribute consistency of response to general functions of mind is to go back to the old innate faculties of the soul. Thus, most behaviorists attempt to analyze all generalization of ability in the individual in terms of a very specific developmental history. As a matter of fact all psychologists who accept the stimulus-response formula must of necessity follow some explanation which makes general ability a combination of specific factors.

The principle of cross-conditioning of E. B. Holt accounts for the consistent and characteristic behavior of the individual in widely varying situations. In regarding such behavior as unrelated to any specific stimuli, we fail to notice that the early habitual activities of the child may become conditioned to constant sources of stimulation both external and internal. This process of cross-conditioning is described by Holt as follows:

> While stooping over his books the boy's total afferent pattern comprises many impulses besides those which are stimulated by the desk and books to which he is immediately responding. There are the characteristic sights, sounds, and odours of the school-room, and also various more or less continuous afferent nerve impulses from the boy's own internal organs. All of these concomitant afferent impulses, according to Pavlov's law of the conditioned reflex, are to some extent acquiring *collateral motor outlet* into the lad's posture of sitting and leaning forward over his books. As this continues day by day, and if it is not counteracted by distinct change of posture and diversity of motor activity, the stooping posture besides being well canalized will come to be so *steadily innervated* by such extraneous and ubiquitous afferent impulses as to be beyond correction. And this is dynamogenic or cross-conditioning. Of these concomitant stimulations, the sights and odours of the school-room will be active only when the boy is there, but all the afferent im-

> pulses from organic sources he carries about within him. And when any posture or motor habit has been so persistently maintained as to become, more deeply than any other postures and habits, cross-conditioned to afferent impulses from the internal organs (as, say, heartbeat and breathing rhythm) it has become in very truth "second nature" (12, 223f.).

Not only may the characteristic posture of the individual be cross-conditioned, but also more complex habits may be similarly maintained by constant stimuli. The child, for example, who is encouraged to declaim and act on every occasion may easily acquire permanent exhibitionistic traits. Then in situations which do not call for acting the child will continue to pose and strut. And the generalized value attitudes of Spranger may be accounted for in a similar manner.

The principle of cross-conditioning thus shows how the characteristic activities of the individual are a function of specific forms of stimulation. As a matter of fact, all stimulus-response theories of development must of necessity involve specific and identical elements. Thus, a consistently good memory is explained by the use of identical techniques in memorizing different types of material. Likewise, a generalized concept of plants is formed through a learning process of selecting out certain features common to all plants.

It should be noted, however, that not all consistency in behavior has as yet been explained on the basis of identical elements. The Gestalt psychologists have raised the very pertinent question of what specific stimulus properties the individual responds to, when he perceives forms and grasps relationships. Even animals can be trained to respond to the brighter colored of two objects in spite of changes in the absolute values of the two brightnesses. Since the absolute values of the stimuli can change without altering the response, the animals evidently are not responding to the specific stimulus of the brighter object. It follows that organisms can respond to the relations in situations with little regard for the specific content of the situation. This ability to respond to the same relation in vary-

ing stimulus settings is regarded as a general function which does not operate through specific identical elements.

The role of cross-conditioning and sustained reflexes in integration. The importance of cross-conditioning in the development of those modes of behavior which characterize an individual's personality has already been mentioned. E. B. Holt points out, however, that cross-conditioning alone will not produce an integrated personality. Holt describes another process essential to integration, namely the development of simultaneously sustained reflexes (12). Movements which are at first random become sustained on the basis of the principle of the reflex circle. Furthermore, these sustained reflexes are directed toward various objects in the environment. Now, the effect of cross-conditioning is to narrow the individual's behavior to given fields of activity. The effect of sustained reflexes, however, is to diffuse the individual's responses to many features of his environment.

The fanatical reformer in politics or the ardent narrow-minded specialist in any field is an individual who has been strongly cross-conditioned, but one who sustains relatively few reflex responses to the general range of objects and events that are immediately about him. He is not alive to these things. At first glance it may appear that such an individual is well integrated, since his responses are organized about a single type of interest. But this organization excludes too much. The individual fails to take cognizance of many things which affect his personal welfare and even of circumstances which bear, not immediately perhaps and yet vitally, if he but knew it, on his specialty itself. He is so concentrated upon matters close to his heart that he is absent-minded and disastrously ineffective in the ordinary traffic of life. His reactions to practical situations (when he reacts at all) are dissociated from the 'whole man,' and are often pathetically inadequate or even childish. Indeed, there is no whole man, and such mutilated central personality as there is gets expression only in dealing with a very narrow section of the environment.

The well-integrated individual, on the other hand, has sufficient cross-conditioning to have interests and a character of his own, but he also possesses many sustained reflexes to the significant objects in his everyday world. He can respond to most situations in life with a great many of his own attitudes and habits. Relatively speaking, such a person continually brings his whole personality into play rather than fragmentary portions of it.

In short we can differentiate among three general types of behavior. The individual who lacks cross-conditioning is poorly integrated. His behavior is primarily a function of the immediate situation. He is all things to all men. The individual who is cross-conditioned but who has relatively few sustained reflexes also lacks true integration. His behavior is primarily a function of himself. He may burn to death in his laboratory, because his absorption in his work prevented a timely discovery of the fire. True integration is found in that personality in whom a balance between sustained reflexes and cross-conditioning has been maintained. His behavior is both a function of himself and of the many situations in life in which he is placed.

The self and personality dissociation. The older psychological descriptions of personality were in terms of the self. The self was described introspectively by William James as that which every individual calls *me* or *mine*. In this description William James included the idea of the social self, the knowledge of oneself as reflected from the attitudes and reactions of other human beings. The self was generally regarded, moreover, not as isolated parts of consciousness but as an essential unity of personality which differentiated the individual from his fellows.

The concept of the self, then, was an introspective term for the integration of personality. The experimental evidence, it will be recalled, is that most individuals show consistency of behavior indicative of integrated activity, but not the perfect consistency indicative of complete integration. Instead of conceiving of an individual's self as representative of all his habits

and attitudes, it is necessary to postulate a number of selves. In most individuals one self will be dominant, but in a minority two or more selves may assert themselves with relatively equal frequency. In the latter case, we speak of the selves as dissociated personalities. It must be remembered, however, that dissociation is a relative term and that most individuals are not at harmony with themselves in all respects. The difference between the dissociated personality and the normal individual is a difference in the degree of integration. Extreme dissociation is often called insanity, but insanity is judged only partly on this basis.

Dissociation of the personality may be simultaneous or successive. An example of simultaneous dissociation is automatic writing which occurs in hysterical patients and under certain circumstances in normal individuals. The subject will write answers to questions, whispered in his ear by one person, while carrying on a conversation with another person. Both his conversation and his written answers are sensible and meaningful. Apparently, two habit systems are functioning in the individual at one and the same time with little or no relation between them.

In successive dissociation, the usual activities of the individual are suddenly broken off and replaced by a different mode of life. The individual frequently has no consciousness of his former self in his new personality role. Often he alternates between his two personalities. In such a dual personality, the individual has two major action systems so incompatible with one another that both cannot be given expression at the same time.

One of the first recorded cases of double personality is that of Mary Reynolds (13). Her friends and relatives knew Miss Reynolds as a reserved, timid, melancholy, even morbid creature. One morning she awoke with all memory of her previous existence gone. She even had to relearn the acts of reading and writing. Her disposition, moreover, had completely changed. She was now fearless, buoyant, and gregarious. After five weeks she lapsed back into her first personality with no knowledge

of what had befallen her. These alternations from one state to another continued at intervals of varying length for sixteen years, finally leaving her in her second state. Gradually, however, the second personality was modified so that in old age it no longer represented a complete emotional opposition to her first state. Some measure of integration had evidently been achieved.

THE CULTURAL DETERMINATION OF PERSONALITY

The determination of personality by culture is a truism of such long standing that social scientists refer to the behavior of the individual as if it were only a miniature reflection of the great society. The significant truth in cultural determination is thus obscured by conceiving of society as an organic whole, imposing its wishes upon its members. This conception makes culture an underived social entity, regards social forms as static and immutable and replaces all explanation of social events by pure tautology. For example, it explains the low rate of homicide in England by asserting that the English are a law-abiding social group.

Society, or *group,* is a shorthand symbol for the fact that many individuals live together in certain relationships to one another. Their habits and attitudes are not permanently and rigidly fixed by a superorganic agency called culture. They are always in the process of change and are the result of a long trial-and-error history of human frustrations and human satisfactions. The significant point, however, is that this process has been going on for so many centuries that the habitual activities of our generation are not the *simple* result of our own trial-and-error learning. The specific teachings of parents and elders and the particular culture objects such as machines, buildings, books, periodicals, newspapers, movies and works of art of our civilization are of basic importance in the development of our personalities. The infant who starts life among Australian bushmen is confronted with radically different social stimulation from the infant born in an American city.

Necessarily the social behavior of the two children will differ markedly. In this sense personality is limited and determined by the social environment.

The conditioning of personality by culture does not mean that the patterns of behavior we learn from our parents are photostatic copies of their habits. They become modified to suit our own purposes and wishes. Some items drop out completely; others are altered, and new items are added. Because the name given to a group of activities remains the same, we are deceived into believing that the pattern of human action has not changed. For example, the names of the major American political parties have remained the same for generations, but the social realities represented by these names have undergone considerable change.

The effects of the cultural environment upon personality can be considered both from the point of view of single personality characteristics and from that of personality integration. Personality characteristics are culturally conditioned in four ways.

In the first place, the rigid institutional and customary practices already in existence among any group of people define in advance for the newcomer the *particular* fields in which he can *not* express his personality. It is true, an individual or individuals may upset this definition, but its existence during his early years is an important factor in his development. What we tend to regard as an inherent natural trait of personality, because it is present in our civilization, may be so institutionalized in another culture as to lose its meaning as a personality characteristic. The wiping out of individual differences in certain fields of behavior through a stereotyping of responses, moreover, may be limited to particular social and economic groups within one culture. Thus in England the lower classes are brought up from birth to be submissive rather than ascendant. Ascendant-submissive behavior is institutionalized or standardized for this group of people.

In the second place, the *number* of fields in which the individual can express his personality is partially determined by the particular culture in which he is reared. It is generally the

aim of dictatorships to institutionalize people on as many items of behavior as possible. Dictators strive for a highly unified and homogeneous group. The citizen becomes identified with the soldier, and the ideal is a nation of like-minded, regimented units with no personality. The citizen-soldier exhibits personality characteristics only when off duty, and, to the extent that he is never allowed off duty, the student of personality finds his field of study vanishing. The avowed ideal of democracies, on the other hand, is the free expression of as many personality characteristics as are possible without mutual self-destruction.

A third way in which personality is culturally affected lies in the extent to which people differ from one another in a given characteristic. In any society, limits to the range of personality variation in many fields are already established in the habits of the people. The base line on which we measure personality differences is generally defined culturally, and we find this base line narrowing or broadening from culture to culture. For example, the range of personality differences in aggressiveness may be much greater among Americans than among Chinese. Within a given culture, of course, the same narrowing or broadening process occurs with the passage of time.

Finally, the culture of a people determines for the developing child the presence or absence of certain personality characteristics, and this apart from the denial of personality expression through institutionalizing forms of behavior. Certain primitive groups living a simple mode of life lack the particular neurotic tendencies of a people in a highly complex civilization. For example, the anthropologist, Malinowski, has observed no hysterical or nervous disorders among the Trobriand Islanders, a Melanesian people of the South Seas (15). It is important, therefore, to distinguish between personality characteristics that arise in any society due to the essential fact of social interaction and characteristics that are the sole result of a particular form of social organization.

The effects of the cultural environment on personality integration can be profitably studied with reference to the com-

plexity of social organization. Although the tendency to oversimplify primitive culture is to be condemned, it is nevertheless true that a society without written records has a much smaller bulk of traditional material and represents a simpler social organization than a culture with a written tradition. This is necessarily so, since a culture without written records is limited by the unaided memories and habits of a relatively small group of people (17).

A complex culture generally represents a type of social organization in which many institutionalized forms are interposed between the individual and the satisfaction of his biological needs. This interposition makes a genuine integration of personality difficult. Contrast, for example, our civilization where the individual satisfies his basic wants indirectly through activities meaningless in themselves with the social organization in the Trobriand Islands where the individual's work is tied up fairly directly to his biological needs and wants. In this Melanesian society, social institutions and customs are organized about an exchange of services and goods between two village communities (15). A coastal village supplies an inland community with fish and in return receives vegetables. Every coastal villager has his partner in the inland community with whom the exchange is made. His personality thus has an opportunity for developing harmoniously in relation to his work and his economic and social contacts.

In our culture the division of labor has been carried to such a fine point that the occupation of most individuals is limited to a narrow field of behavior. Personality may be affected in two ways. We may become narrow specialists and center our interests in our specialty so that we fail to express our personalities in many of the situations in which we are placed. Or we may develop no character of our own and respond to situations without regard for the way in which that response fits in with the rest of our behavior. The dilemma here is that the complexity which narrows personality presents it with a wider social field to which to adjust.

The opportunities for personality integration in a complex

culture are also determined by the relationships existing between its dominant institutions. In a society where social, political and economic institutions are in essential conflict, personalities are necessarily at war with themselves. In a civilization where government, industry, religion and art are harmoniously related, personalities tend to be integrated. It is precisely at this point that psychology can help most in our problems of social living. If we are to live happily, our social institutions must not be arranged from the standpoint of whether they are good institutions in themselves, but they must receive their sanctions according as they make for the welfare and happiness of human beings.

REFERENCES

1. Allport, F. H. *Social psychology.* Cambridge: Houghton Mifflin Co., 1924.
2. Allport, F. H. and Hartman, D. A. The measurement and motivation of atypical opinion in a certain group. *Amer. Pol. Sci. Rev.,* 1925, 19, 735-760.
3. Allport, F. H., Schanck, R. L., and Dickens, M. Psychology in relation to social and political problems. In *Psychology at work* (Achilles, P. S., editor). New York: McGraw-Hill Co., 1932. Chapter 7.
4. Allport, G. W. A test for ascendance-submission. *J. Abn. & Soc. Psychol.,* 1928, 23, 118-136.
5. Allport, G. W., and Vernon, P. E. The field of personality. *Psychol. Bull.,* 1930, 27, 677-730.
6. Allport, G. W., and Vernon, P. E. *Studies in expressive movement.* New York: Macmillan Co., 1933.
7. Cantril, H., and Allport, G. W. Recent applications of the *Study of Values. J. Abn. & Soc. Psychol.,* 1933, 28, 259-273.
8. Freeman, F. N., Holzinger, K. J., and Mitchell, B. C. The influence of environment on the intelligence, school achievement, and conduct of foster children. *27th Yrbk. Nat. Soc. Educ.,* 1928. Pt. 1, 103-218.
9. Fryer, D. *The measurement of interests.* New York: Henry Holt and Co., 1931.
10. Guilford, J. P., and Braly, K. Extroversion and introversion. *Psychol. Bull* 1930, 27, 96-107.
11. Hanna, J. V. Clinical procedure as a method of validating a measure of psychoneurotic tendency. *J. Abn. & Soc. Psychol.,* 1934, 28, 435-445.
12. Holt, E. B. *Animal drive and the learning process.* New York: Henry Holt and Co., 1931.

13. James, W. *Principles of psychology.* New York: Henry Holt and Co., 1890.
14. Katz, D., and Allport, F. H. *Students' attitudes.* Syracuse: Craftsman, 1931.
15. Malinowski, B. *Sex and repression in savage society.* New York: Harcourt, Brace and Co., 1927.
16. Maslow, A. H. Dominance and social behavior of monkeys. *Psychol. Bull.,* 1934, 31, 688.
17. Mead, M. The use of primitive material in the study of personality. *Character & Personality,* 1934, 3, 3-16.
18. Newman, H. H. Mental and physical traits of identical twins reared apart. *J. Hered.,* 1932, 23, 3-18.
19. Schanck, R. L. A study of a community and its groups and institutions conceived of as behaviors of individuals. *Psychol. Monog.,* 1932, 43, No. 195.
20. Spranger, E. *Types of men.* Halle: Niemeyer, 1928.
21. Strong, E. K. *Change of interests with age.* Stanford University Press, 1931.
22. Vernon, P. E., and Allport, G. W. A test for personal values. *J. Abn. & Soc. Psychol.,* 1931, 26, 231-248.
23. Vetter, G. B. The measurement of social and political attitudes and the related personality factors. *J. Abn. & Soc. Psychol.,* 1930, 25, 149-189.

INDEX